Е. К. МЕРКУРЬЕВА

ОСНОВЫ БИОМЕТРИИ

ИЗДАТЕЛЬСТВО МГУ

МИНИСТЕРСТВО ВЫСШЕГО И СРЕДНЕГО СПЕЦИАЛЬНОГО ОБРАЗОВАНИЯ РСФСР

НАУЧНО-МЕТОДИЧЕСКИЙ КАБИНЕТ ПО ЗАОЧНОМУ И ВЕЧЕРНЕМУ ОБУЧЕНИЮ МОСКОВСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА ИМЕНИ М. В. ЛОМОНОСОВА

Е. К. МЕРКУРЬЕВА

ОСНОВЫ БИОМЕТРИИ

УЧЕБНОЕ ПОСОБИЕ

для биологических факультетов
государственных университетов

ИЗДАТЕЛЬСТВО МОСКОВСКОГО УНИВЕРСИТЕТА
—1963—

Утверждено
кафедрой генетики и селекции
биолого-почвенного факультета
МГУ

ПРЕДИСЛОВИЕ

Развитие биологической науки требует все более широкого использования разнообразных методов исследования, заимствованных из области точных наук. В связи с этим должна быть усилена математическая подготовка биологов, в частности, и по такому ее разделу, как вариационная статистика.

Использование и освоение имеющейся литературы по вариационной статистике представляет значительную трудность для большинства студентов-биологов и особенно для тех из них, которые обучаются на вечерних и заочных отделениях университетов.

В помощь студентам-биологам и составлено настоящее пособие, в основу которого положен лекционный материал и материал практических занятий, проводившихся автором в течение 12 лет при чтении курса биометрии на кафедре генетики и селекции Московского государственного университета имени М. В. Ломоносова для студентов, специализирующихся в области генетики и селекции животных и растений. Этим объясняется то обстоятельство, что большинство примеров дается из области генетики и селекции, и в них рассматриваются объекты, являющиеся предметом исследования кафедры. Для облегчения самостоятельной работы студентов-заочников с другими пособиями и книгами по вариационной статистике в тексте используются также примеры, заимствованные у других авторов.

В данном пособии не ставится цель изложить теорию вариационной статистики и дать вывод статистических формул и их математическое обоснование. Весь текст направлен на то, чтобы студент научился приемам обработки биологического материала вариационно-статистическим методом и имел при подготовке своих курсовых и дипломной работ справочный материал в виде статистических формул и подсобных математических таблиц.

Данное учебное пособие может быть полезно для студентов и специалистов, обучающихся и работающих в области сельскохозяйственных и медицинских наук.

Выражаю благодарность профессору А. Я. Боярскому и научному сотруднику А. В. Пугачевой за просмотр рукописи и ценные указания.

ГЛАВА I

ПРЕДМЕТ И МЕТОД ВАРИАЦИОННОЙ СТАТИСТИКИ. МЕСТО МАТЕМАТИЧЕСКОГО МЕТОДА В БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Закономерности многих явлений в природе и обществе очень часто могут быть вскрыты только тогда, когда они изучаются не на единичных случаях или фактах, или объектах, а на массовом материале, т. е. при многократном повторении опытов и при постоянных условиях, в которых проводится испытание.

Предположим, что требуется установить длительность эмбрионального периода у животных какого-либо вида или породы. Для этого необходимо собрать массовый материал, характеризующий этот признак при определенных условиях, т. е. получить сведения на многих объектах данного вида или породы. Так, например, на массовом материале установлено, что длительность эмбрионального развития у кролика равна 30 дням, но часть самок вынашивает детенышей 29 дней, часть — 30 и какая-то часть самок имеет длительность беременности 31 день.

Возьмем другой пример: для выявления нормального уровня температуры тела человека необходимо провести обследование многих людей. На массовом материале установлено, что норма этого показателя равна 36,6° С. Но при измерении температуры тела у многих лиц обнаруживаются отклонения от этой нормы: можно встретить здоровых людей с температурой тела в 37,0° С, 36,2°, 36,5° С и т. п. И только в массе, или, как говорят, в среднем, выявляется нормальный уровень температуры тела, равный 36,6°, при значительном разнообразии (варьировании) его у разных лиц.

В ы я в л е н и е   з а к о н о м е р н о с т е й   р а з л и ч н ы х процессов или объектов на массовом материале при наличии изменчивости (вариабильности) исследуемого признака, возникающей при определенном комплексе условий в результате многих случайных причин, и составляет задачу вариационной статистики.

Предметом вариационной статистики является варьирующий признак, а массовые наблюдения объектов (совокупности объектов) и обобщающие их результаты (показатели) составляют ее метод.

Такое содержание предмета и метода вариационной статистики означает, что теоретическим фундаментом этой науки является теория вероятностей и, в частности, закон больших чисел.

В общей форме закон больших чисел утверждает, что основные закономерности какого-либо массового процесса проявляются в обобщающих показателях, в которых исчезает элемент случайного, если в них объединяются единичные явления или факты в достаточно большом количестве.

В работах К. Маркса неоднократно указывается на роль больших чисел для правильного понимания и анализа закона стоимости в капиталистическом обществе. В. И. Ленин много раз подчеркивал значение закона больших чисел для выявления экономических закономерностей, критикуя буржуазных экономистов за неправильные выводы, сделанные на основе единичных фактов.

В разработке теоретических основ вариационной статистики большое значение имеют исследования русской школы математиков. В 1846 г. вышла работа выдающегося русского математика М. В. Остроградского, в которой автор впервые сделал попытку использовать математический метод случайной выборки при изучении промышленного производства.

Развитие этого метода в дальнейшем осуществляли такие крупные отечественные ученые, как П. Л. Чебышев, А. М. Ляпунов, А. А. Марков, именами которых назван ряд теорем в области теории вероятностей. В последние годы советская школа математиков в лице ученых А. Н. Колмогорова, С. Н. Бернштейна, А. Я. Хинчина, В. И. Романовского, Б. В. Гнеденко, Н. В. Смирнова, Б. С. Ястремского, Е. Е. Слуцкого и других продолжила разработку проблем вариационной статистики и создала ряд учебных пособий общетеоретического и прикладного характера.

Использование математического метода в биологических исследованиях, начатое бельгийским антропологом Кетле (1796—1874), позднее нашло применение в работах Гальтона, Пирсона, но их идеалистическое мировоззрение сказалось на их работах. За последние годы существенный вклад сделан английским математиком Р. А. Фишером, работающим в области сельскохозяйственных и биологических проблем.

С развитием биологических наук методы вариационной статистики все шире применяются для изучения разнообразных биологических процессов Так, генетик Иогансен использовал этот метод при обработке материалов, полученных им в опытах на чистых линиях фасоли, ботаник Клебс изучал с помощью биометрии изменчивость растений по числу тычинок и по другим признакам под влиянием внешних условий, ботаник Бонне выяснял изменчивость одуванчика при измененных условиях жизни, генетики Дорфмейстер, Вейсман, Штандфусс, Фишер использовали математический анализ при изучении сезонного диморфизма в окраске бабочек под влиянием изменения температуры окружающей среды

Независимо от правильности или неправильности идей того или иного из этих авторов привлечение ими статистических методов соответствует особенностям биологических объектов, как объектов, которым свойственна вариация многих признаков

Статистический анализ, примененный в биологических исследованиях, помог выявить роль внешних условий в изменчивости животных и растительных организмов

Но наряду с положительной стороной использования математического метода для изучения биологических процессов в последние десятилетия наметилась и отрицательная сторона в использовании приемов этого метода, проявляющаяся у представителей буржуазной науки и заключающаяся в том, что в математической статистике все больше и больше перестают видеть ее подсобный характер применительно к биологическим исследованиям и превращают ее в основной метод биологии При этом упускается специфика биологических явлений и недооценивается изучение их качественной природы биологическими методами Чисто формальный подход без учета качественной природы биологических объектов особенно виден в ряде генетических работ, исходящих из корпускулярной теории наследственности, которая придает большое значение случайным процессам, например, таким, как случайная комбинаторика хромосомного аппарата, перекомбинация элементов ДНК, случайное слияние мужских и женских гамет в процессе оплодотворения и т. п.

Правильное использование математических приемов для биологии исходит из того, что они должны быть подчиненными, а не ведущими методами биологических исследований.

Необходимо остановиться и на научном значении единичных наблюдений в биологии. С точки зрения статистического подхода, единичные наблюдения не представляют научной ценности. Иное дело, если оценивать роль единичных наблюдений и фактов по сути их биологического значения. Известны многие примеры, когда единичные наблюдения имеют большое научно-познавательное значение. Так, например, И. В. Мичурин методом вегетативной гибридизации получил гибрид груши с яблоней «Ренет бергамотный». Единственный экземпляр этого вегетативного гибрида послужил источником селекционного материала и был использован для выведения новых сортов плодовых. Кроме того, факт получения отдаленного гибрида вегетативным методом имел большое теоретическое значение.

Значение для науки единичных наблюдений особенно хорошо видно на палеонтологических находках, когда по единственному найденному экземпляру ископаемого делается обобщение об особенностях вида в целом.

Таким образом, использование математического метода при изучении массовых биологических материалов должно базироваться на анализе специфической природы биологического объекта, сочетаться с применением для познания его свойств биологических методов исследования и не исключать бережного и внимательного отношения к единичным фактам и явлениям.

С помощью вариационной статистики можно изучать количественные и качественные показатели различных явлений и объектов.

Вариационная статистика, примененная к биологическим объектам, выделилась в специфический раздел и получила название биометрии, введенное в 1899 г. Гальтоном.

ГЛАВА II

ОСНОВНЫЕ СТАТИСТИЧЕСКИЕ ПОНЯТИЯ И ЗАДАЧИ ВАРИАЦИОННОЙ СТАТИСТИКИ

Биологические явления, как и другие явления природы и общества, характеризуются массовостью. Массовые материалы могут быть получены или в специально поставленных опытах, или путем обследования и наблюдения. Так, например, для изучения влияния облучения на физиологическое состояние животных производится закладка специальных опытов, для чего выделяется определенное число подопытных животных, из которых комплектуются подопытные и контрольные группы. Примером сбора массовых наблюдений путем обследования могут служить энтомологические работы в поле, направленные на выявление степени зараженности растений вредителем.

Собранные тем или иным путем данные массовых наблюдений подвергаются статистической обработке и анализу, на основании которых могут быть легче выяснены биологические закономерности, характерные для изучаемого показателя.

Выше указывалось, что **предметом вариационной статистики являются варьирующие признаки**. Разберем, что составляет суть и особенности варьирующих признаков.

Варьирующими признаками могут служить многие показатели, например, такие, как живой вес, рост, химический состав, окраска, степень пораженности вредителем и т. п.

Так, если взять для изучения в качестве показателя высоту растений одного сорта, выращенных на одной делянке при одновременном сроке высева всех зерен, то окажется, что, несмотря на сортовую однородность изучаемого объекта и на одинаковые условия произрастания, индивидуальные показатели роста растений, измеренных в один и тот же срок, оказываются неодинаковыми. Одни растения будут очень малы, часть растений будет иметь высокий рост, а остальные распределятся определенным образом в границах роста между самыми низкими и самыми высокими растениями.

Иначе говоря, в пределах однородной группы и при сходных условиях опыта или наблюдения величина изучаемого признака неодинакова у разных особей, входящих в группу обследования, или, как принято говорить, признак варьирует, т. е. проявляет изменчивость.

Варьирование степени выраженности признака у объектов, составляющих однородную группу, при определенных сходных условиях обусловлено воздействием на объект многих действующих случайных причин или факторов. Одни из них могут вызвать у данной особи увеличение признака, другие, наоборот, его уменьшение, что в конечном счете обусловливает определенную величину развития признака как у каждого единичного объекта, так и у всех остальных объектов, вошедших в изучаемую однородную группу.

Так, например, известно, что плодовитость кролика породы шиншилла выражается у одних самок числом рожденных детенышей, равным 2, у других — 5, 6, 7 или 8, у некоторых плодовитость доходит до 12 крольчат в помете; средняя же плодовитость равна 8. На показатель плодовитости самок могут влиять условия эмбрионального питания для каждого эмбриона, качество родительских гамет, из которых образовались зародыши, и многие другие неизвестные нам факторы.

При этом оказывается, что число животных в обследованной группе, имеющих очень низкую или очень высокую плодовитость, будет значительно меньше, чем число животных, имеющих плодовитость, близкую к средней. То есть, частота встречаемости особей с различным отклонением уровня варьирующего признака от среднего его значения различна.

Закономерности варьирования признака и распределение объектов по его величине в пределах однородной группы массового наблюдения вскрываются на основе ряда теорем теории вероятностей [1].

Разберем понятие вероятности, с которым постоянно имеет дело вариационная статистика, на следующем примере: если при инкубации выводится 45% петушков (p) и 55% куро-

---

[1] См. Б. В. Гнеденко и А. Я. Хинчин. Элементарное введение в теорию вероятностей. Физматгиз, М., 1961

чек (q), то говорят, что вероятность получения птицы мужского пола равна 0,45 (или 45%).

Определение вероятности формулируется так: вероятность P события A равна отношению числа p исходов, благоприятствующих появлению этого события, к общему числу n всех возможных несовместимых и равновозможных исходов испытания.

«Вероятность есть мера объективной возможности данного события... Вероятность этого события измеряется... отношением числа благоприятствующих шансов к общему числу шансов» [2].

Вероятность события выражается дробью, величина которой находится в пределах от 0 до 1, что можно представить так:

$$0 \leqslant \frac{p}{n} \leqslant 1.$$

Чем больше значение дроби $\frac{p}{n}$, тем чаще наступает событие A. Если вероятность события близка к нулю, то это событие наступает крайне редко или совершенно не наступает. Если же вероятность события близка к единице, то это означает, что событие наступает очень часто (почти всегда).

Следовательно, чем меньше вероятность, тем реже может встречаться данное событие или объект, особь с данной величиной варьирующего признака и, наоборот, чем больше вероятность, тем чаще встречается данное событие.

Так, например, если известно, что рождение мужских особей имеет вероятность, равную 0,6, то из 150 рожденных детей вероятнее всего, что родится 90 мальчиков и 60 девочек.

Следует различать понятия «вероятность» и «частость». Вероятность как мера возможности реализации случайного события A проявляется только при очень большом числе наблюдений или, как говорят, при большом числе повторений опытов. Так, например, известно, что при большом числе наблюдений соотношение полов (мужского и женского) у новорожденных детей близко 1 : 1. Следовательно, вероятность рождения мальчика или девочки одинакова и составляет 0,5. Но если проверить это соотношение на малом числе новорожденных, например, в пределах одной семьи, то часто оно меняется в ту или другую сторону. Так, из 10 детей в одной семье могут родиться 3 мальчика и 7 дево-

[2] А. М. Длин. Математическая статистика в технике. Изд-во «Сов. наука», М., 1958.

чек или может иметь место и еще более резкое расхождение в соотношении полов. Если из 10 детей родилось 3 мальчика, то дробь $^3/_{10}$ не будет служить показателем вероятности, а называется частостью (или относительной частостью).

Частость, так же, как и вероятность, изображается дробью и изменяется в пределах от 0 до 1. При увеличении числа наблюдений частость стремится к вероятности.

Следовательно, в конкретных условиях суждение о частоте появления события можно получить, исходя из величины частости, которая может быть вычислена на основании данных, полученных в опыте или собранных путем обследования или наблюдения.

Величина же вероятности появления интересующего нас события проявляется при осуществлении большого количества массовых наблюдений и поэтому ее можно назвать идеальной мерой возможности появления этого события.

Так как частость и вероятность между собой связаны, то по частости можно судить приближенно о вероятности и, наоборот, по вероятности можно заранее предсказывать частоту появления какого-либо события.

Так, например, теоретическое расщепление, т. е. соотношение особей с доминантным и рецессивным признаками во втором поколении гибридов, по правилам Менделя, дает вероятность появления особей с доминантным признаком, равную 0,75. Следовательно, если в конкретных условиях предполагается иметь в опыте 1000 гибридов, то из них вероятнее всего, что у 750 будет доминантный признак, а у 250 — рецессивный.

Для варьирующих признаков характерно «распределение вероятностей». Так, вероятность появления объектов (или особей), имеющих величину варьирующего признака, близкую к минимальной или максимальной границам варьирования, очень мала, в то время, как объекты, обладающие уровнем варьирующего признака, близким к среднему значению, имеют более высокое значение вероятности и, следовательно, их появление, встречаемость или численность будут больше.

В самом деле, всем известно, как редко встречаются люди, имеющие очень высокий или очень низкий рост и, наоборот, чаще всего встречаются люди среднего роста или близкие к нему.

Из сказанного видно, что величина варьирующего признака, изменяющегося под влиянием случайных причин, служащего предметом изучения вариационной статистики, есть величина переменная, которая может принимать разные значения с определенной вероятностью. Такие величины называются случайными.

В статистике принято варьирующий признак или, как мы уже сказали, случайную переменную величину изображать буквой *x*. В большинстве же пособий по вариационной статистике для биологов варьирующий признак обозначается буквой *V*. Варьирующий признак называют еще вариантом.

Варьирующий признак может быть как количественным и измеряться любой мерой (сантиметрами, килограммами, микронами и др.), так и качественным (окраска, форма, пол, тип заболевания, физиологическое состояние, генерация и т. п.). Как количественные, так и качественные случайные величины могут служить предметом статистической обработки при массовых наблюдениях.

Варьирующие признаки разделяются на непрерывные и прерывные. Непрерывные варьирующие признаки (например, рост растения, живой вес животного, удой) могут принимать любое промежуточное значение в границах его варьирования, в том числе и дробное. Прерывные признаки (число зёрен, число яиц) могут принимать только целые значения в границах варьирования.

Статистические величины какой-либо группы наблюдений образуют так называемый статистический ряд, пределами которого ограничивается вся изменчивость данной группы объектов.

Пусть имеется подопытная группа, состоящая из десяти самок кролика, на которой изучается такой варьирующий признак, как плодовитость. Если выписать в строчку (или в виде столбика) данные о плодовитости каждой крольчихи, то тем самым будет составлен статистический ряд, состоящий из десяти наблюдений (частот). Если обозначить варьирующий признак или варианты через *x*, то статистический ряд будет записан так:

№ животного	1	2	3	4	5	6	7	8	9	10
Варьирующий признак, x	6	8	10	11	12	14	15	14	10	12

Самой простой характеристикой этого ряда служит степень изменчивости признака и его среднее значение.

Так, размах изменчивости плодовитости самок в данной подопытной группе определяется показателями минимальной и максимальной плодовитости и ограничен показателями плодовитости от 6 до 15 крольчат. Среднюю плодовитость, т. е. среднее значение варьирующего признака, можно установить здесь простым способом, а именно, сложив все значения вариантов и разделив эту сумму на число животных в группе. Частное от такого деления и будет служить показателем средней арифметической, обозначаемой через букву M, или $\bar{x}$.

Формула средней арифметической выглядит так:

$$\bar{x} = M = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\Sigma x}{n} = \frac{112}{10} = 11{,}2 \text{ крольчонка.}$$

Из приведенного выше видно, что даже самый простой статистический ряд имеет свои статистические характеристики.

Статистические ряды называют также **вариационными рядами**. Они могут быть образованы как из показателей количественных признаков, так и из показателей, отличающихся качественно.

Примером ряда, составленного из показателей качественных признаков, может служить ряд по особенностям шёрстного покрова кроликов. Шёрстный покров кроликов может быть короткошёрстным или пуховым, пигментированным или альбиносным. Предположим, что имеется группа помесных кроликов, которые имеют следующий тип шёрстного покрова:

№ кролика	1	2	3	4
Варьирующий признак, x	Белый пуховый	Белый короткошёрстный	Черный пуховый	Черный короткошёрстный

Следующее необходимое понятие — **статистическая совокупность**, которая представляет собой обобщение статистических рядов, когда в них производится объединение в один класс особей с одинаковым или близким значениями величин варьирующего признака.

В указанном выше примере с плодовитостью самок кролика статистическая совокупность выглядит так:

Класс (вариант) по плодовитости, x	6	7	8	9	10	11	12	13	14	15
Число самок в каждом классе (частоты), p . .	1	0	1	0	2	1	2	0	2	1

Общее число наблюдений (т. е. число самок) в этой совокупности равно $n = \Sigma p = 10$ (буквой p обозначается число объектов в каждом классе, которое получило название «частот»).

Примером статистической совокупности может служить распределение колхозов по уровню средней урожайности с гектара (табл. 1).

Таблица 1

Статистический ряд распределения колхозов по средней урожайности ржи с 1 *га* в *ц*

Класс по урожайности, x	10—14	15—19	20—24	25—29
Число колхозов (частоты), p	2	15	30	5

Статистический ряд для качественных признаков представлен в табл. 2 на примере с типами шёрстного покрова кроликов.

Таблица 2

Статистический ряд распределения кроликов по типам шёрстного покрова

Класс, x	Белые пуховые	Черные пуховые	Белые коротко-шёрстные	Черные коротко-шерстные
Число голов (частоты), p	5	20	15	10

Любой статистический ряд может быть подвергнут статистической обработке, на основании которой определяется несколько характеристик, отражающих особенности и закономерности этого ряда.

Большое значение в вариационной статистике имеют такие два понятия, как генеральная совокупность и случайная выборка.

Положим, требуется определить всхожесть зерна, засыпанного в элеватор. Совокупность всех зёрен на элеваторе будет представлять генеральную совокупность. Чтобы определить всхожесть, характеризующую зерно этой генеральной совокупности, подвергают контрольному проращиванию зёрна, взятые в виде проб из различных мест элеватора. Полученные данные о всхожести зёрен из этих проб будут выборочным наблюдением, и такие пробы называются выборкой, или выборочной совокупностью.

По характеристикам выборочной совокупности, полученной в результате статистической обработки, можно судить о генеральной совокупности. Иначе говоря, выборочная совокупность представляет часть генеральной и по этой части судят о целом, т. е. о генеральной совокупности.

Выборочная совокупность должна правильно отражать генеральную совокупность или, как говорят, она должна быть репрезентативной, т е. представительной.

Выборочная совокупность значительно уменьшает объём статистической работы, так как выборка всегда имеет меньшее число наблюдений (или объектов), чем генеральная совокупность, и в то же время выборка позволяет иметь правильное суждение о генеральной совокупности.

Изучение непосредственно генеральной совокупности в ряде случаев нельзя осуществлять потому, что оно должно было бы сопровождаться уничтожением объектов, составляющих её. Возможность использования выборочного материала устраняет это препятствие. В самом деле, если требуется узнать содержание гемоглобина или эритроцитов в крови, то для проведения анализа не требуется использовать всю кровь данного объекта, а достаточно определить эти показатели в капле крови, проба которой должна правильно отражать среднее качество крови у изучаемого животного. Приведенный ранее пример с выборочным определением всхожести зёрен из проб, взятых на элеваторе, также указывает на большое значение выборки в изучении генеральной совокупности, так как для определения всхожести достаточно подвергнуть проращиванию зёрна выборочных проб.

Что же нужно сделать для того, чтобы выборка была репрезентативной и правильно отражала, или представляла генеральную совокупность?

Для репрезентативности выборки обязательными условиями являются присутствие в выборке представителей всех классов варьирующего признака и пропорциональность численности их в каждом классе соответствующей численности каждого класса генеральной совокупности.

Поясним это на примере. На поле озимой ржи требуется взять для апробации образцы растений. При этом интересующим нас признаком служит число стеблей в кусте одного растения, характерное следующим варьированием: один, два, три, четыре и пять стеблей — это будут показатели классов. Предположим, что в генеральной совокупности растения распределяются по этому признаку так:

Число стеблей в кусте растения, x . .	1	2	3	4	5
Число растений в % от общего количества (частоты), p	10	20	35	20	15

Если поставлена задача взять на данном поле выборочно 1000 растений, удовлетворяя при этом требованию, что в выборке должны быть представлены растения всех классов пропорционально их отношению в генеральной совокупности, то требовалось бы включить в выборку 100 одностебельных растений, 200 — двухстебельных, 350 — трехстебельных, 200 — четырехстебельных и 150 — пятистебельных. Выделенные таким образом растения могут обрабатываться статистически по признаку кустистости. Но при этом возникает большая трудность, а именно: мы обычно не знаем, как варьирует изучаемый признак в генеральной совокупности и каково распределение растений по классам варьирующего признака.

Для преодоления указанной трудности прибегают к такому приёму, как получение так называемой случайной выборки, основывающемуся на теории вероятностей и законе больших чисел. Отбор объектов в случайную выборку из генеральной совокупности производится по принципу жеребьёвки, когда включение той или иной особи в выборку осуществляется независимо от желания исследователя, и каждый член генеральной совокупности имеет возможность (или шанс) попасть в случайную выборку.

Полученная на основе этих принципов достаточно большая случайная выборка будет правильно представлять генеральную совокупность, т. е. она будет репрезентативной.

Для получения случайных выборок существует несколько приёмов, а именно:

1. **Выборка с повторением.** Представим, что объекты (члены) генеральной совокупности нумеруются по порядку, и эти номера записываются на талончики. Затем талончики поме-

щаются в урну, перемешиваются и из них методом жеребьёвки отбирают нужное число талончиков, которые должны образовать случайную выборку. Если талон после вытаскивания возвращается в урну, то выборка называется повторной.

2. **Бесповторная выборка** получается тем же способом, что и повторная, только вынутый талон не возвращается в урну к остальным талонам.

На основании номеров, записанных на попавших в выборку талонах, дальнейшему изучению будут подвергаться те члены генеральной совокупности, номера которых вошли в случайную выборку.

Этот приём должен осуществляться как для отбора объектов в случайную выборку при обследованиях, так и для отбора объектов в специальные эксперименты.

Разберем пример получения случайной выборки. Положим, что в экспериментальном саду имеется 1000 сеянцев яблонь сорта антоновка «шестисотграммовая». В задачу исследования входит изучение прироста растений в первый год жизни. Для этих целей намечено произвести обследование 30% всех сеянцев, т. е. сделать выборку объёмом в 300 растений. Так как сеянцы размещены на участке в определенном порядке и имеют индивидуальные номера, то составляют 1000 талонов с указанием на каждом талоне индивидуального номера сеянца. Талоны помещают в урну, перемешивают и по методу повторной или бесповторной выборки вытаскивают 300 талонов. Номера, написанные на вынутых талонах, будут указывать, что именно тот или иной сеянец должен быть обследован по интересующему нас признаку.

Принцип случайности в составлении выборки должен сохраняться и для тех случаев, когда нельзя осуществить нумерацию членов генеральной совокупности, например, при отборе для анализа проб крови, образцов зерна и т. п.

Принцип случайности при этом осуществляется путем взятия проб зёрен из разных мест элеватора или проб растений в шахматном порядке с полевого участка, путем взятия крови из периферических сосудов в виде типичной капли, путем взятия средней пробы кашицы из мышечной ткани для производства биохимического анализа и т. п.

Если выборка имеет субъективно организованный или преднамеренный характер, то обработка ее данных вариационно-статистическим методом теряет свой смысл.

Существенным моментом при составлении выборки является однородность и однотипность условий, в которых находятся члены генеральной совокупности.

Под качественной однородностью выборки понимается то, что она составлена из объектов, близких по своим существенным свойствам или, как говорят, по общим признакам.

Так, например, если изучается живой вес скота, то в выборку могут включаться только животные одной породы, одного пола и близкие по возрасту. Пол, порода и возраст в данном случае являются общими признаками, а живой вес, подвергающийся изучению, называется основным признаком.

При составлении выборки желательно, чтобы объекты ее были качественно однородны по нескольким общим признакам.

Для изучения основных признаков выбираются наиболее важные, которые и будут служить предметом статистического анализа. Такими признаками в нашем последнем примере, кроме живого веса, могут быть показатели удоя, жирномолочности, скороспелости и т. п.

Для получения однородности в случайной выборке необходимо, чтобы и генеральная совокупность отвечала этим же качествам.

Под однородностью условий подразумевается, что члены (объекты) генеральной совокупности, намеченные к изучению, находятся в близких или сходных условиях. Для биологических объектов это означает, что изучение вопроса должно происходить на фоне общих кормовых, климатических, почвенных и других факторов, которые в случае невыравненности их для отдельных членов генеральной совокупности могут привести к неправильному анализу и выводу.

Таким образом, еще до того, как исследователь приступает к получению выборки, он должен знать свой материал по указанным признакам и обеспечить однородность объектов и условий по общим показателям.

Выборка, имеющая объём численностью до 30 вошедших в неё объектов ($n \leqslant 30$), называется малой выборкой, и для неё имеются свои методы обработки и требования.

Выборка, имеющая объём численностью свыше 30 объектов ($n > 30$), называется большой выборкой. Позднее мы разберем вопрос о том, какое число наблюдений требуется включить в опыт, чтобы получить статистически достоверные характеристики изучаемого показателя.

После того, как составлена случайная выборка, интересующие нас показатели членов этой выборки подвергаются статистической обработке.

Статистическая обработка материала позволяет:

1. Определить средние характеристики варьирующего при-

знака и, тем самым, выявить типические черты изучаемых объектов, сгруппированных в однородную совокупность.

2. Найти величину, форму и тип варьирования изучаемого признака и установить закономерности распределения частот для объектов с различным значением варьирующего признака.

3. Вычислить величину и направление связи между двумя или большим числом признаков.

4. Определить статистическую достоверность проведенных опытов или собранных наблюдений и учесть величину ошибок различных статистических характеристик, вычисленных по материалам выборки.

5. Сопоставить эмпирические данные с теоретическими.

6. Применить графический метод анализа статистических переменных величин.

7. Выявить долю влияния отдельных факторов в общей изменчивости варьирующего признака.

Решение этих задач осуществляется различными приёмами обработки массовых данных с применением различных формул, вытекающих из теоретических положений вариационной статистики.

ГЛАВА III

СРЕДНИЕ ВЕЛИЧИНЫ И ИХ СВОЙСТВА. МОМЕНТЫ И ИХ ИСПОЛЬЗОВАНИЕ ПРИ ВЫЧИСЛЕНИЯХ

При изучении какого-либо варьирующего признака в исследуемой группе объектов мы вынуждены прибегать к определению средних показателей, которые служат обобщающими характеристиками интересующего нас признака у группы объектов и отражают типические особенности этой группы.

В зависимости от того, какая задача ставится при вычислении характеристик данной группы обследованных объектов, можно пользоваться различными средними показателями, каждый из которых имеет свое математическое значение и смысл.

Необходимо иметь в виду, что все средние имеют абстрактное значение. Так, при расчетах можно получить дробное значение показателя (например, средняя плодовитость кролика 6,5 крольчонка), который в конкретных условиях не может иметь места.

Существуют следующие основные средние:

средняя арифметическая — M или $\bar{x}$;
средняя геометрическая — G
средняя квадратическая — S;
мода — Mo;
медиана — Me.

1. СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ M ИЛИ $\bar{x}$

Наиболее употребляемая средняя — это средняя арифметическая M, которую обозначают также через $\bar{x}$.

Для вычисления средней арифметической требуется сложить все показатели варьирующего признака x, и полученную сумму разделить на число наблюдений n, т. е. $M = \frac{\Sigma x}{n}$.

Пример. Определить средний вес тёлочек при рождении:

№ животного	1	2	3	4	5	6	7	8	9	10
Вес в *кг* — варьирующий признак, x	30	35	40	38	30	28	33	30	31	34

$$\Sigma x = 329;$$

$$\bar{x} = M = \frac{\Sigma x}{n} = \frac{329}{10} = 32{,}9 \text{ кг}.$$

Таким образом, средняя арифметическая есть частное от деления суммы варьирующего признака на число членов случайной выборки (или, как говорят, на число наблюдений).

Основные свойства средней арифметической

1. Алгебраическая сумма отклонений показателя варьирующего признака каждого члена выборки от средней арифметической этого вариационного ряда всегда равна нулю, т. е. $\Sigma(x-\bar{x})=0$. Рассмотрим это свойство средней арифметической на примере с живым весом новорожденных тёлок (стр. 22). Средняя арифметическая этого ряда равна 32,9 *кг*.

Подсчитаем отклонение живого веса для каждой тёлки от этой средней, т. е. определим $x - M$.

Вес тёлки в *кг*, x	30	35	40	38	30	28	33	30	31	34
Средняя арифметическая, M	32,9	32,9	32,9	32,9	32,9	32,9	32,9	32,9	32,9	32,9
Отклонение в *кг*, $(x - M)$	−2,9	+2,1	+7,1	+5,1	−2,9	−4,9	+0,1	−2,9	−1,9	+1,1

Алгебраическая сумма положительных и отрицательных отклонений равна: $-15{,}5+15{,}5=0$.

Следовательно, $\Sigma(x-M)=0$.

2. Сумма квадратов отклонений от средней арифметической всегда меньше суммы квадратов отклонений от любого другого числа a, т. е.:

$$\Sigma(x-M)^2 < \Sigma(x-a)^2.$$

Продемонстрируем это свойство на том же примере. Возьмём за произвольное число a величину живого веса в 30 *кг* и вычислим отклонения показателей живого веса от этого числа $(x-a)$ для каждого животного. Возведём в квадрат числа, полученные как отклонения от средней арифметической $(x-\overline{x})^2$, и числа, полученные путём отклонения от условного числа $(x-a)^2$:

№ телки	1	2	3	4	5	6	7	8	9	10
Ряд отклонений от $\overline{x}$, $(x-\overline{x})$	−2,9	+2,1	+7,1	+5,1	−2,9	−4,9	+0,1	−2,9	−1,9	+1,1
$(x-\overline{x})^2$	8,41	4,41	50,41	26,01	8,41	24,01	0,01	8,41	3,61	1,21
Ряд отклонений от условного числа a, $(x-a)$	0	5	10	8	0	−2	3	0	1	4
$(x-a)^2$	0	25	100	64	0	4	9	0	1	16

Сумма квадратов отклонений будет следующей:

$$\Sigma(x-\overline{x})^2 = 134{,}9;$$

$$\Sigma(x-a)^2 = 219{,}0.$$

Следовательно, сумма квадратов отклонений от средней арифметической x равна 134,9, а от условной величины a она оказалась значительно больше, равной 219,0.

2. ВЗВЕШЕННАЯ СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ $\overline{x}_{взв}$

Кроме простой средней арифметической зачастую необходимо вычислить взвешенную среднюю арифметическую, так как использование простой средней может дать неправильный показатель.

Так, например, если требуется определить средний урожай в пяти колхозах, то необходимо вычислить среднюю взвешенную. Под «взвешиванием» в статистике понимают умножение варианта x на число наблюдений (или частоту) p, соответствующих данному варианту.

Следовательно,

$$\overline{x} = \frac{x_1p_1 + x_2p_2 + \ldots + x_np_n}{p_1 + p_2 + \ldots + p_n} = \frac{\Sigma xp}{\Sigma p}.$$

Пример. Определить средневзвешенный показатель урожайности озимой пшеницы в пяти колхозах района:

№ колхозного хозяйства	1	2	3	4	5
Средняя урожайность зерна с 1 *га* в *ц*, x	15	12	18	20	10
Посевная площадь ржи в *га*, p	600	900	1000	700	800
xp	9000	10800	18000	14000	8000

Если определить простую среднюю арифметическую, то из этих данных было бы получено следующее:

$$\overline{x} = \frac{\Sigma x}{n} = \frac{75}{5} = 15\,ц,$$

где n — число хозяйств.

Если же вычислить взвешенную среднюю арифметическую по формуле $\overline{x} = \frac{\Sigma xp}{\Sigma p}$, то получим правильную величину среднего урожая по группе пяти колхозов, а именно:

$$\overline{x} = \frac{\Sigma xp}{\Sigma p} = \frac{59800}{4000} = 14{,}95\,ц.$$

Чем больше было бы различие в урожайности и посевных площадях этих хозяйств, тем больше отличались бы величины простой и взвешенной средних.

По такому же принципу, например, следует определять у коров средний процент жира за лактационный период, что видно из следующего примера:

Месяц лактации	1	2	3	4	5	6	7	8	9	10
% жира, x	3,0	3,0	3,0	3,2	3,2	3,4	3,5	3,7	4,0	4,2
Удой в *кг*, p	400	600	400	360	300	260	200	160	100	60
xp	1200	1800	1200	1152	960	884	700	592	400	252

Простая средняя, по этим данным, $\overline{x} = \frac{\Sigma x}{n} = \frac{34{,}2}{10} = 3{,}42\%$, взвешенная средняя $\overline{x} = \frac{\Sigma xp}{\Sigma p} = \frac{9140}{2840} \approx 3{,}22\%$. Как видно, обе средних существенно отличаются друг от друга.

3. СРЕДНЯЯ ГЕОМЕТРИЧЕСКАЯ G

Средняя геометрическая вычисляется по следующей формуле:

$$G = \sqrt[n]{x_1 \cdot x_2 \ldots x_n},$$

где n — число наблюдений в вариационном ряду;
x_1, x_2 и т. д. — показатели варьирующего признака.

Логарифмируя это выражение, получаем:

$$\lg G = \frac{\Sigma \lg x}{n} = \frac{\lg x_1 + \lg x_2 + \ldots + \lg x_n}{n}.$$

Так, если необходимо определить среднюю геометрическую вариационного ряда, характеризующегося такими вариантами x: 5,8,9,4,3, то $G = \sqrt[5]{5 \cdot 8 \cdot 9 \cdot 4 \cdot 3}$. После логарифмирования получаем:

$$\lg G = \frac{\lg 5 + \lg 8 + \lg 9 + \lg 4 + \lg 3}{5} =$$

$$= \frac{0,6990 + 0,9031 + 0,9542 + 0,6021 + 0,4771}{5} = \frac{3,6355}{5} = 0,7271,$$

откуда $G = 5,34$.

Средняя геометрическая применяется в тех случаях, когда требуется определить средний относительный прирост (темп) какого-либо показателя или явления за определенный отрезок времени. Для этих целей используется следующая формула средней геометрической:

$$G = \sqrt[n]{j_1 \cdot j_2 \cdots j_n},$$

где j — есть отношение последующего показателя к предыдущему;
n — число периодов, на протяжении которых определяется прирост.

Так, если требуется определить среднегодовой прирост живого веса у телят за первый год жизни, то следует вычислить среднюю геометрическую.

Предположим, живой вес телят джерсейской породы изменялся следующим образом (табл. 3):

Таблица 3

Статистический ряд распределения телят по живому весу

Возраст в месяцах	При рождении	3	6	9	12
Живой вес в *кг* . . .	20	73	135	190	235
i		$\frac{73}{20}=3,65$	$\frac{135}{73}=1,85$	$\frac{190}{135}=1,40$	$\frac{235}{190}=1,23$

Отсюда средняя геометрическая годичного прироста равна:

$$G=\sqrt[4]{3,65\cdot 1,85\cdot 1,40\cdot 1,23}=\sqrt[4]{11,63}=1,85 \text{ (или } 185\%).$$

Средняя геометрическая может быть вычислена путем логарифмирования по формуле:

$$G=\sqrt[n]{\frac{y_1}{y_2}},$$

где n — число периодов; y_1 — уровень показателя на конечный срок; y_2 — на начальный срок. В нашем примере $G=\sqrt[4]{\frac{235}{20}}$. Логарифмируя эти данные, получаем:

$$\lg G=\frac{\lg 235-\lg 20}{4}=\frac{2,3711-1,3010}{4}=\frac{1,0701}{4}=0,2675,$$

откуда $G=1,85$. Таким образом, вычисление средней геометрической по обеим формулам дает одинаковые результаты.

4. СРЕДНЯЯ КВАДРАТИЧЕСКАЯ S

Средняя квадратическая используется в ряде случаев при вычислении степени изменчивости признака. Ее вычисление производится по следующей формуле:

$$S=\sqrt{\frac{\Sigma x^2}{n}}.$$

Следовательно, средняя квадратическая представляет собой корень квадратный из частного, полученного от деления суммы квадратов значений вариантов на число наблюдений, вошедших в выборку.

Проведем на следующем примере сопоставление средних величин $\bar{x}$, S, G, характеризующих вариационный ряд.

Пример. Определить средние величины $\bar{x}$, S, G для вариационного ряда, составленного из показателей плодовитости белых мышей после облучения:

№ самки	1	2	3	4	5
Число мышат в помете, x	3	5	1	2	4
x^2	9	25	1	4	16

$$\bar{x} = \frac{\Sigma x}{n} = \frac{15}{5} = 3 \text{ головы},$$

$$S = \sqrt{\frac{\Sigma x^2}{n}} = \sqrt{\frac{55}{5}} = \sqrt{11} = 3{,}317 \text{ головы},$$

$$G = \sqrt[n]{x_1 \cdot x_2 \ldots x_n} = \sqrt[5]{3 \cdot 5 \cdot 1 \cdot 2 \cdot 4} = \sqrt[5]{120}.$$

Логарифмируя это выражение, получаем:

$$\lg G = \frac{\Sigma \lg x}{n} = \frac{\lg 3 + \lg 5 + \lg 1 + \lg 2 + \lg 4}{5} =$$

$$= \frac{0{,}4771 + 0{,}6990 + 0{,}0 + 0{,}3010 + 0{,}6021}{5} = \frac{2{,}0792}{5} = 0{,}4158,$$

откуда $G = 2{,}6$ головы.

Таким образом, полученные арифметическая, квадратическая и геометрическая средние для данного ряда имеют следующие соотношения между собой и между минимальным и максимальным значениями вариантов:

$$\text{минимальный вариант } x = 1 < G = 2{,}6 < \bar{x} = 3 < S = 3{,}317 < \begin{array}{c}\text{максимальный}\\ \text{вариант } x = 5\end{array}$$

Из написанного видно, что все средние величины занимают срединное место между минимальным и максимальным значениями варьирующего признака: средняя геометрическая G

имеет значение, меньше средней арифметической $\bar{x}$, средняя квадратическая S превышает среднюю арифметическую $\bar{x}$.

5. МОДА Mo

Часто в качестве средней характеристики вариационных рядов служит мода.

Модой называется чаще всего встречающееся значение варианта. Это означает, что модальный вариант имеет наибольшую частость.

Использование моды особенно удобно, если изучаются качественные признаки (форма, окраска, тип шёрстного покрова).

Примером модального варианта может служить следующий: если из 215 помесных кур у 35 желтая окраска ног, у 120 — аспидная, а у 60 — зелёная, то аспидная окраска в данном примере — модальный признак, так как этот тип пигментации чаще встречается чем другие.

Если выборка состоит из большого числа наблюдений, распределенных по классам вариационного ряда (см. стр. 28), то класс с наибольшим числом наблюдений называется модальным, и в его пределах находится значение Mo, которое можно вычислить по следующей формуле:

$$Mo = x_{M_o} + k\frac{(p_2 - p_1)}{(2p_2 - p_1 - p_3)},$$

где x_{Mo} — начало модального класса;

k — величина класса;

p_1 — частота класса, предшествующего модальному;

p_2 — частота модального класса;

p_3 — частота класса, следующего за модальным.

Пример. Определить значение моды в ряду распределения коров джерсейской породы по живому весу:

Класс варьирующего признака по весу в *кг*, x	250—299	300—349	350—399	400—449	450—499	500—549
Число наблюдений (частоты), p	4	15 p_1	30 p_2	10 p_3	5	1

$x_{Mo} = 350$ *кг*;
$k = 50$ *кг*;
$p_1 = 15$;
$p_2 = 30$;
$p_3 = 10$.

Подставляем значения вариационного ряда в формулу:

$$Mo = x_{Mo} + k \frac{(p_2 - p_1)}{(2p_2 - p_1 - p_3)} = 350 + 50 \frac{(30 - 15)}{(2 \cdot 30 - 15 - 10)} =$$

$$= 350 + 50 \cdot \frac{15}{35} = 350 + 21,4 = 371,4 \ кг.$$

Величина живого веса коров джерсейской породы, равная 371,4 *кг,* и будет служить показателем моды; следовательно, чаще всего будут встречаться животные, имеющие вес, близкий к значению моды.

В зависимости от характера вариационного ряда значение моды может быть близким к средней арифметической, могут быть и такие ряды, когда значение моды смещено в сторону максимального или минимального значений вариантов, и даже совпадает с ними. Встречаются вариационные ряды, в которых имеют место две моды. В дальнейшем мы разберем такие специфические вариационные ряды.

6. МЕДИАНА ИЛИ СРЕДИННОЕ ЗНАЧЕНИЕ ВАРИАЦИОННОГО РЯДА *Me*

Медианой называется значение того варианта, который делит всю совокупность наблюдений, расположенную в порядке возрастания или убывания значений вариантов, на две равные части.

Использование медианы удобно при изучении качественных признаков.

Величина медианы при большом числе наблюдений вычисляется по следующей формуле:

$$Me = x_{Me} + k \frac{(i_1 - i_2)}{p_{M_e}},$$

где x_{Me} — начало класса, в котором находится медиана;

k — величина класса;

i_1 — число вариантов, соответствующее половине всех наблюдений;

p_{Me} — частота медианного класса;

i_2 — число вариантов в классах, предшествующих медианному.

Пример. Найти значение медианы для ряда распределения кустов помидор по числу плодов на каждом кусте:

№ куста	1	2	3	4	5	6	7	8	9	10	11
Число плодов	14	15	14	17	13	20	16	22	13	18	14

Для определения медианы при малом числе наблюдений располагают все варианты в возрастающем порядке. При этом данные нашего примера будут размещаться следующим образом:

13 13 14 14 14 **15** 16 17 18 20 22.

Вариантом, выражающим медиану, будет шестой вариант этого ряда, равный 15 плодам.

Пример. Определить медиану вариационного ряда живого веса коров, приведенного на стр. 30.

При обработке этого ряда для вычисления медианы требуется применить приём, который получил название «накопленные частоты»:

Класс по живому весу в *кг*, x	250—299	300—349	350—399	400—449	450—499	500—549
Частоты, p	4	15	30	10	5	1
Накопленные частоты	4	19	49	59	64	65

Величина класса $k=50$ *кг*.

Ряд накопленных частот, изображенный на стр. 30 в виде третьей строчки, составляется путем последовательного сложения частоты каждого последующего класса с частотой предыдущего; при этом полученная сумма проставляется в каждый класс. Так, в нашем примере в первом классе (250—299 *кг*) частота равна 4. К ней прибавляется частота следующего класса (300—349 *кг*), выражающаяся числом 15; от сложения 4 и 15 получаем сумму 19, которая и будет суммой накопленных частот для второго класса. К этой сумме прибавляется затем частота следующего класса: именно: $19+30=49$, что служит показателем накопленных частот третьего класса. Продолжая дальше это суммирование, доходим до последнего класса, где накопленные частоты составляют число 65. Дальнейшее вычисление ведется по формуле, приведенной на стр. 29.

Так как медиана представляет собой вариант, делящий всю совокупность пополам, то доля вариантов, приходящихся на каждую из сторон от медианы, составляет 50% и выражается через $q_l=0{,}5$. Отсюда, если q_l умножить на число наблюдений n, то полученное число вариантов i_1 образует половину всех наблюдений, вошедших в вариационный ряд, т. е. $i_1=q_l\cdot n$.

В приведенном примере $i_1=q_l\cdot n=0{,}5\cdot 65=32{,}5$.

После отыскания значения i_1 находим класс, в пределах которого находится значение медианы. Для этих целей используется ряд накопленных частот. Отыскиваем класс, в котором сумма накопленных частот превышает значение i_1, равное 32,5. Таким классом в нашем примере является третий класс по счету, имеющий значение накопленных частот, равное 49. Этот третий класс и заключает в своих границах значение медианы. Начало третьего класса — 350 *кг* — в нашем примере служит показателем x_{Me}, входящим в формулу. Найдем значение i_2 по ряду накопленных частот. Сумма накопленных частот, полученная в третьем классе, т. е. предшествующем медианному, и есть значение i_2. В нашем примере $i_2=19$.

Подставляем все необходимые данные в формулу для вычисления медианы:

$$Me=x_{Me}+k\cdot\frac{(i_1-i_2)}{p_{Me}}=350+50\cdot\frac{(32{,}5-19)}{30}=$$

$$=350+\frac{5\cdot 13{,}5}{3}=350+22{,}5=372{,}5\ \textit{кг}.$$

Таким образом, в нашем примере серединное (центральное) значение будет иметь медианный вариант, величина которого оказалась равной 372,5 *кг*.

Модой же этого вариационного ряда, как это было вычислено (см. стр. 29), служит вариант, равный 371,4 *кг*.

Подсчитаем для этого же ряда значение средней арифметической M. Для этого следует перемножить серединное значение каждого класса на соответствующее ему значение частоты p; полученные произведения xp сложить и разделить на число наблюдений, т. е. $M=\frac{\Sigma xp}{n}$. Для получения серединного значения класса следует к значению варианта, стоящего в начале класса x, прибавить половину классового промежутка, т. е. $\frac{k}{2}=\frac{50}{2}=25$ *кг*.

В нашем примере:

Середина класса по живому весу в *кг*, x	275	325	375	425	475	525
Частоты, p	4	15	30	10	5	1
xp	1100	4875	11250	4250	2375	525

Отсюда средний живой вес

$$M = \frac{\Sigma xp}{\Sigma p} = \frac{24375}{65} = 375{,}0 \text{ кг}.$$

Сопоставим между собой полученные среднюю арифметическую, моду и медиану для данного вариационного ряда. Они оказались следующими: $M=375{,}0$ *кг*; $Mo=371{,}4$ *кг*; $Me=372{,}5$ *кг*.

Соотношение этих средних, а также средней геометрической и средней квадратической, обычно оказывается следующим:

минимальный вариант x	<	средняя геометрическая G	<	средняя арифметическая, мода, медиана (M, Mo, Me)	<	средняя квадратическая S	<	максимальный вариант x

7. МОМЕНТЫ СТАТИСТИЧЕСКИХ ВЕЛИЧИН

Моментом статистической величины называется сумма произведений тех или иных степеней отклонений варианта x от средней арифметической M или от условной величины A на соответствующую частоту p, деленная на сумму всех частот (Σp или n).

Это определение может быть представлено в общей форме в виде следующего выражения:

$$\frac{\Sigma (x-M)^n \cdot p_n}{n} \text{ или } \frac{\Sigma (x-A)^n \cdot p_n}{n}.$$

Использование моментов основано на свойствах средней арифметической и на свойствах отклонений вариантов от средней арифметической, о чем указывалось на стр. 22 и 23.

Моменты статистических величин могут быть начальными (обозначаются через m) и центральными (обозначаются через μ).

Моменты различаются между собой степенью, в которую возводится отклонение суммы вариантов от средней арифметической или от условной средней.

Для формул различных статистических характеристик используются моменты разных степеней, что и будет показано ниже.

Техника вычисления моментов, а следовательно, и техника вычисления тех статистических характеристик, в формулу которых они входят, может быть различной. Можно в этих целях пользоваться методом произведений, методом сумм и методом разностей. В наших последующих примерах мы будем пользоваться методом произведений и методом сумм.

8. ТЕХНИКА ОБРАБОТКИ ВАРИАЦИОННЫХ РЯДОВ

Обработка случайной выборки начинается с составления вариационного ряда. При большом числе наблюдений составляется вариационный ряд с разбивкой всех вариантов по классам. После этого применяется тот или иной метод обработки частот вариационного ряда для вычисления интересующих нас статистических характеристик.

Разберем на примере технику составления вариационных рядов и вычисление средних величин.

Предположим, что произведено обследование стада коров джерсейской породы в количестве 104 голов по показателю живого веса. Первичным материалом для вариационных вычислений служит опись животных с указанием их веса. Имея указанную опись, следует начать обработку этих данных с определения величины классового промежутка и составления классов вариационного ряда. Для этого находим крайние показатели живого веса в выборке и определяем размах изменчивости. Просматривая данные описи, выясняем, что максимальный вес обследованных животных оказался равным 500 *кг* и минимальный — 320 *кг*. Определяем разницу между этими крайними вариантами:

$$\begin{array}{r} \text{максимальный вес} = 500 \textit{ кг} \\ \underline{\text{минимальный вес} = 320 \textit{ кг}} \\ \text{разница } D = 180 \textit{ кг} \end{array}$$

Зная разницу D, найдем величину класса, которая будет служить мерой разбивки всех членов выборочной совокупности на классы вариационного ряда. Считая, что удобным количеством классов в вариационном ряду будет 10, делим разницу на 10 и получаем величину классового промежутка k:

$$k = \frac{D}{10} = \frac{180}{10} = 18 \textit{ кг}.$$

Для удобства составления вариационного ряда в примере величину класса k можно округлить до 20 *кг*

После этого составляем границы классов Границы первого класса должны быть такими, чтобы в него вошел вариант с минимальным значением (320 *кг*) Последний класс должен иметь границы, в которые войдет максимальный вес (500 *кг*). Минимальный и максимальный варианты могут быть как на границе класса, так и внутри него

На основе полученной величины $k=20$ *кг* составляем классы, беря за границы первого класса значение варианта от 320 до 339 *кг* При этом второй класс будет иметь границы от 340 до 359 *кг* и т д (стр 35)

Считается удобным иметь следующее число классов в зависимости от числа наблюдений в выборке при n от 40 до 60 рекомендуется иметь 6—8 классов, при 61—100 наблюдениях — 7—10 классов, при 100—200 наблюдениях — 9—12 классов, при 200—500 наблюдениях — 12—17 классов

После составления классов вариационного ряда производится разноска имеющихся в описи показателей вариантов по классам

Положим, что в описи коров первое животное имеет вес 395 *кг,* следовательно, оно должно быть отнесено к четвертому по счету классу, второе животное имеет вес 450 *кг* и относится к седьмому классу Занося сведения о весе каждого животного в соответствующий класс, мы ставим против этого класса точку, затем, когда в классе стоят уже четыре точки в виде квадратика , соединяем линиями стороны квадратика при каждом занесении варианта , после чего вписываем диагонали и получаем такую фигуру $\boxtimes$. Такой квадратик будет означать, что в этот класс занесено десять наблюдений После подсчета числа наблюдений в каждом классе получаем ряд частот p Сумма частот по классам (Σp) должна быть равна числу наблюдений (n)

Таким образом, разноска данных позволила получить вариационный ряд или, как еще говорят, ряд распределения

Распределение частот в вариационном ряду имеет определенную закономерность, заметную даже на первый взгляд, а именно число коров с большим и малым живым весом незначительно и, наоборот, в середине вариационного ряда накапливается большое число наблюдений В этих особенностях вариационного ряда проявляется варьирование случайных величин той генеральной совокупности (в нашем случае варьирование живого веса коров джерсейской породы), которая представлена случайной выборкой

Варианты классов, x	Частоты, p	Условные отклонения, a	pa
320—339	4	—3	—12
340—359	11	—2	—22
360—379	24	—1	—24
A 380—399	22	0	0
400—419	20	+1	20
420—439	13	+2	26
440—459	5	+3	15
460—479	2	+4	8
480—499	2	+5	10
500—519	1	+6	6
	$\Sigma p = n = 104$		$\Sigma pa = +27$

Для обработки полученного вариационного ряда, вычисления моментов и статистических характеристик воспользуемся методом произведений.

Для этого выделим класс с так называемой условной средней A, которая будет равна середине этого класса (желательно, чтобы она была близка к средней арифметической). Таким классом обычно выбирают класс, имеющий большее число наблюдений и расположенный в середине вариационного ряда. В нашем примере этим классом может быть взят класс, имеющий границы 380—399 *кг* и значение частот 22, но может быть выбран и следующий за ним класс.

Отчеркнем линиями строчку этого класса и примем его за нулевой Тогда вверх от него пойдут классы с уменьшающимися значениями вариантов. Их отклонение от нулевого класса обозначается порядковыми числами со знаком минус, т. е. —1, —2, —3. Вниз от нулевого класса варьирующий признак увеличивается, и отклонения классов перечисляются порядковыми числами со знаком плюс, т. е. +1, +2, +3, +4, +5, +6.

Этот ряд отклонений классов от нулевого (т. е. от класса с условной средней A), выраженный порядковыми числами, обозначаем через a (условное отклонение).

Для вычисления средней арифметической M данного ряда следует пользоваться формулой: $M = A + k \cdot m_1$, в которую введен первый начальный момент $m_1 = \frac{\Sigma pa}{n}$.

Таким образом, средняя арифметическая имеет формулу:

$$M = A + k \cdot \frac{\Sigma pa}{n},$$

где A — вариант условной средней;
n — число наблюдений;
k — величина класса;
Σpa — сумма произведений частот вариационного ряда на соответствующее значение условного отклонения a.

Проделаем необходимые расчеты в нашем примере: перемножим частоту p каждого класса на его условное отклонение a и просуммируем полученные произведения pa, учитывая знаки. Получаем: $\Sigma pa = -58 + 85 = +27$.

Для получения условной средней A, которая является серединой класса, нужно к нижней границе этого класса прибавить половину классового промежутка k, т. е. $380 + \frac{20}{2} = 390$ *кг*, — эта величина и будет значением условной средней A.

Подставляя необходимые величины в формулу средней арифметической, вычисляем ее значение:

$$M = A + k \cdot \frac{\Sigma pa}{n} = 390 + 20 \cdot \frac{27}{104} = 390 + 5{,}19 = 395{,}19 \text{ кг}.$$

Для вычисления средней арифметической была использована формула первого начального момента m_1. Для ряда других статистических характеристик используются моменты второго, третьего и четвертого порядков, о чем будет сказано ниже.

Формулы же этих начальных моментов таковы:

$$m_2 = \frac{\Sigma pa^2}{n}; \quad m_3 = \frac{\Sigma pa^3}{n}; \quad m_4 = \frac{\Sigma pa^4}{n}.$$

Из этого видно, что последовательное умножение частот на соответствующую степень условного отклонения a^2, a^3, a^4 позволяет вычислить все указанные моменты без особой затраты времени на технику вычисления.

В ряде случаев для удобства расчетов и для упрощения формул различных статистических характеристик используются центральные моменты μ, которые могут быть выражены через начальные моменты m следующим образом:

$$\mu_0 = 1; \ \mu_1 = 0; \ \mu_2 = m_2 - m_1^2; \ \mu_3 = m_3 - 3m_2 \cdot m_1 + 2m_1^3.$$

Использование этих формул будет показано в ходе дальнейшего изложения.

ГЛАВА IV

ИЗМЕНЧИВОСТЬ И МЕТОДЫ ЕЕ ИЗМЕРЕНИЯ

В биологии большое значение имеет изучение явления изменчивости. Методы изучения этого явления могут быть различными, специфичными для каждой отрасли биологической науки. Так, например, в генетике и селекции такими специфическими приёмами будут: скрещивание, вегетативная гибридизация, стадийный анализ, половой ментор и пр. Для изучения биологических явлений могут быть привлечены и использованы методы других биологических наук, а также методы, взятые из физики, химии, математики. Так, например, изучение обмена веществ и характера ассимиляции может осуществляться физиологическими и биохимическими методами; используя радиоактивное излучение, можно вызывать изменчивость наследственности организмов; методом математической статистики можно выяснить особенности индивидуальной изменчивости (вариабильности) животных и растений, с которой приходится сталкиваться постоянно, работая с видом, сортом, породой, стадом, опытной делянкой или группой объектов.

В этой главе дается ознакомление с математическими методами измерения степени индивидуальной изменчивости.

Математический метод изучения изменчивости не вскрывает причин этого явления, а лишь констатирует степень вариабильности (изменчивости) интересующего нас признака в группе объектов, вошедших в случайную выборку и отражающих свойства генеральной совокупности. Этот метод может выявить некоторые особенности в типе изменчивости.

Статистический метод позволяет выявить изменчивость, возникающую у организмов под влиянием многих факторов, действие которых может осуществляться в разных направлениях: одни — увеличивают, другие — уменьшают величину признака или показателя, и в результате совместного действия таких часто нами неучтенных факторов формируется та величина признака у каждого объекта, с которой мы и сталкиваемся при его изучении. Но следует помнить, что математический метод не может вскрыть истинные причины изменчивости, которые можно обнаружить только с помощью биологических методов исследования.

1. ОПРЕДЕЛЕНИЕ СТЕПЕНИ ИЗМЕНЧИВОСТИ МЕТОДОМ ЛИМИТОВ И СРЕДНИМ КВАДРАТИЧЕСКИМ ОТКЛОНЕНИЕМ σ

Простыми показателями степени изменчивости служат лимиты, или размах изменчивости, определяемый по разнице между максимальным и минимальным значениями признака.

В разбиравшемся выше примере (стр. 33) изменчивость живого веса у коров джерсейской породы выражается лимитом от 500 до 320 *кг*, т. е. 180 кг. Если в другом стаде лимит по весу животных равен 100 *кг*, то можно сделать заключение, что во втором стаде изменчивость этого показателя ниже, и стадо более константно по этому признаку.

Этот способ выражения степени изменчивости применяется и при публикации научных данных, но он очень примитивен.

Для выражения степени изменчивости используются квадраты отклонений каждого варианта от средней арифметической M (или от условной средней A). Корень квадратный из их средней величины стали называть средним квадратическим отклонением или стандартным отклонением и изображать буквой σ (сигма)[3]:

В общем виде это выражается так:

$$\sigma = \sqrt{\frac{\Sigma (x - \overline{x})^2}{n}} \text{ или } \sigma = k\sqrt{\mu_2}.$$

Если рассматривать *отклонения* от условной средней A, то, используя начальные моменты, получаем: $\sigma = k\sqrt{m_2 - m_1^2}$, что означает:

$$\sigma = k\sqrt{\frac{\Sigma pa^2}{n} - \left(\frac{\Sigma pa}{n}\right)^2}.$$

---

[3] В ряде пособий среднее квадратическое отклонение обозначается латинской буквой *S*.

Квадрат условного отклонения, т. е. σ^2 (или S^2) также служит мерой изменчивости и называется дисперсией (см. гл. VIII).

Среднее квадратическое отклонение показывает, насколько в среднем каждый вариант статистической совокупности отклоняется от средней арифметической.

Среднее квадратическое отклонение σ — величина именованная, она имеет то же наименование, что и изучаемый признак (*кг, м* и т. п.). Весь лимит изменчивости (от максимума до минимума) в нормальных рядах распределения укладывается в интервале $\pm 3\sigma$, что будет подробнее разбираться дальше.

Продолжим обработку нашей выборки и её статистического ряда по живому весу коров джерсейской породы для нахождения σ как показателя степени изменчивости этого признака, пользуясь последней формулой.

Для этого дополним расчёты, сделанные на стр. 35, вычислением ряда pa^2. Перемножим числа столбика pa еще раз на условное отклонение a, т. е. $(-12)\cdot(-3)=36$ и т. д. Произведение pa^2 оказалось для всех классов положительным и $\Sigma pa^2=339$.

x	p	a	pa	pa^2
320—339	4	—3	—12	36
340—359	11	—2	—22	44
360—379	24	—1	—24	24
A 380—399	22	0	0	0
400—419	20	1	20	20
420—439	13	2	26	52
440—459	5	3	15	45
460—479	2	4	8	32
480—499	2	5	10	50
500—519	1	6	6	36
	$\Sigma p = n = 104$		$\Sigma pa = +27$	$\Sigma pa^2 = 339$

Вычислим квадрат первоначального момента m_1^2:

$$m_1^2 = \left(\frac{\Sigma pa}{n}\right)^2 = \left(\frac{27}{104}\right)^2 = 0{,}0676.$$

Подставляем в формулу среднего квадратического отклонения полученные при обработке ряда значения:

$$\sigma = k\sqrt{\frac{\Sigma pa^2}{n} - \left(\frac{\Sigma pa}{n}\right)^2} = 20\sqrt{\frac{339}{104} - \left(\frac{27}{104}\right)^2} =$$

$$= 20\sqrt{3{,}26 - 0{,}0676} = 20\sqrt{3{,}1924} = 20 \cdot 1{,}786 = 35{,}72 \text{ кг.}$$

Таким образом, среднее квадратическое отклонение для данного ряда оказалось равным 35,72 *кг*.

Как было сказано выше, изменчивость в нормальных вариационных рядах ограничивается значением минимального и максимального вариантов, отклоняющихся от средней арифметической на $\pm 3\sigma$.

Пример. Средняя арифметическая по живому весу, полученная при обработке ряда, была равна 395,19 *кг* (стр. 36).

Рассчитаем значение минимального варианта, беря его отклонение от M на -3σ:

$$x_{\text{мин}} = M - 3\sigma = 395{,}19 - 3 \cdot 35{,}72 = 395{,}19 - 107{,}16 =$$

$$= 288{,}03 \text{ кг.}$$

По этому же принципу найдем значение максимального варианта:

$$x_{\text{макс}} = M + 3\sigma = 395{,}19 + 107{,}16 = 502{,}35 \text{ кг.}$$

Фактическое же значение минимального варианта по данным нашей выборки равно 320 *кг*, а максимального — 500 *кг*. Это расхождение теоретического лимита с фактическим объясняется тем, что фактическое распределение не вполне совпадает с нормальным, или же тем, что в пределах $\pm 3\sigma$ от средней арифметической находятся не все 100%, а лишь 99,7% всех особей выборки: 0,3% особей выходят за пределы изменчивости, ограниченной в пределах $\pm 3\sigma$.

Таким образом, среднее квадратическое отклонение есть мера рассеивания вариантов около средней арифметической. Если взять величину среднего квадратического отклонения в квадрате, то это так же может служить мерой рассеивания вариантов, называясь варианцой или дисперсией (σ^2). Вопросу об использовании этой статистической характеристики посвящается специальная глава «Дисперсионный анализ», которая входит в раздел вариационной статистики и позволяет существенно углубить анализ вариации, возникающей под влиянием разнообразных факторов, а также выделить в общей изменчивости долю частной изменчивости, возникающей под влиянием какого-либо конкретного одного или нескольких факторов.

2. ПОПРАВКА ШЕППАРДА НА ВНУТРИКЛАССОВЫЕ КОЛЕБАНИЯ

При группировке членов случайной выборки по классам допускается некоторая неточность в значении среднего квадратического отклонения σ. Это происходит потому, что варианты, вошедшие в границы одного класса, различны, а мы при обработке как бы приравниваем их к величине варианта, приходящегося на середину класса, т. е. усредняем; тем самым игнорируется внутриклассовое колебание вариантов.

Внутри классов варианты распределяются не равномерно, больше в той части, которая ближе к средней, что уменьшает общую колеблемость. Чтобы учесть это, рекомендуется в ряде случаев вводить поправку Шеппарда на внутриклассовое колебание вариантов.

Для этого из σ^2 вычитается постоянная величина, равная $^1/_{12}$ величины класса или 0,083333, умноженная на k^2. Следовательно, корректированное значение σ получается из следующего выражения $\sigma^* = \sqrt{\sigma^2 - k^2\,0{,}0833}$.

3. ВЫЧИСЛЕНИЕ σ МЕТОДОМ СУММ

Воспользуемся тем же вариационным рядом с показателем живого веса коров и обработаем его методом сумм.

Метод сумм заключается в том, что из частот, расположенных по обе стороны от нулевого класса (класс условной средней A), составляется сумма накопленных частот по тому же принципу, как это было показано при вычислении медианы (стр. 30).

Накопление ведется от минимального класса до нулевого и от максимального до нулевого. При этом составляется два ряда накопленных частот, из которых данные первого ряда будут использованы для вычисления первого начального момента m_1 и данные второго — для вычисления второго начального момента m_2. Проведем обработку указанного вариационного ряда методом сумм:

x	p	$q_1 = 58$	$q_3 = 23$
320—339	4	4	4
340—359	11	15	19
360—379	24	39	—
A 380—399	22		
400—419	20	43	—
420—439	13	23	42
440—459	5	10	19
460—479	2	5	9
480—499	2	3	4
500—519	1	1	1
	$\Sigma p = 104$	$q_2 = 85$	$q_4 = 75$

Обозначим верхнюю часть первого ряда накопленных частот через q_1, а нижнюю — через q_2; для второго ряда соответственно — q_3 и q_4.

Ряд q_1 составляется следующим образом: в строчке минимального класса (т. е. 320—339 *кг*) в колонке q_1 записывается величина частот этого класса, равная 4; под этой цифрой записываются в следующем классе накопленные частоты, полученные от сложения частот первых двух классов, т. е. $4+11=15$; в третьем классе накопленные частоты равны: $15+24=39$. По аналогии с этим составляется ряд q_3, где первая цифра повторяет частоту минимального класса; во втором классе накопленные частоты образуются из суммирования накопленных частот первого ряда q_1, т. е. $4+15=19$, а в классе, граничащем с нулевым, делается прочерк.

Для накопления частот от класса с максимальным вариантом делается аналогичный расчет, но только он осуществляется от максимального класса в сторону нулевого. В этом случае первый ряд накопленных частот q_2 будет осуществляться так: 1; $1+2=3$; $3+2=5$; $5+5=10$; $10+13=23$; $23+20=43$. Для второго ряда q_4 накопление частот получается следующее: 1; $1+3=4$; $4+5=9$; $9+10=19$; $19+23=42$. В классе, граничащем с нулевым, делается прочерк.

После накопления частот производится их суммирование: $q_1=4+15+39=58$ и $q_3=4+19=23$, что и записывается сверху столбиков q_1 и q_3; $q_2 = 1+3+5+10+23+43=85$ и $q_4 = 1+4+9+19+42=75$, что записывается внизу столбиков q_2 и q_4.

Если сопоставить эти суммы с теми, которые получены на стр. 39 при обработке этого же ряда методом произведений, то получается, что эти данные, вычисленные разными способами, равны.

Так, например, $q_1=\Sigma[p(-a)]$, т. е. -58;

$q_2=\Sigma[p(+a)]$, т. е. $+85$.

Откуда $\Sigma pa=q_2-q_1=85-58=27$.

Для того, чтобы получить Σpa^2, нужно сложить $q_1+q_2+2(q_3+q_4)=58+85+2(23+75)=339$, что соответствует Σpa^2, полученной методом произведений.

Таким образом, обработка вариационного ряда методом произведений (с использованием условного отклонения a) и методом сумм (с использованием приема накопленных частот) дает одинаковые результаты в вычислении рабочих показателей, используемых в формулах статистических характеристик.

Перейдем к разбору способов вычисления среднего квадратического отклонения σ для малых выборок, т. е., когда число наблюдений меньше 30.

В зависимости от того, с какими вариантами приходится работать, техника расчетов может быть разной.

4. ВЫЧИСЛЕНИЕ σ ДЛЯ МАЛЫХ ВЫБОРОК ПРИ ОДНОЗНАЧНОМ ИЛИ ДВУЗНАЧНОМ ЗНАЧЕНИЯХ ВАРИАНТОВ

Для вычисления σ при однозначных значениях варианта удобно использовать возведение в квадрат значения варианта. Формула среднего квадратического отклонения при этом будет выглядеть так:

$$\sigma=\sqrt{\frac{\alpha}{n-1}}\text{, где } \alpha=\Sigma x^2-\frac{(\Sigma x)^2}{n}.$$

Пример. Определить изменчивость плодовитости подопытной группы свиней, если число наблюдений равно 10.

№ животного	Плодовитость, x	x^2
1	10	100
2	12	144
3	11	121
4	5	25
5	8	64
6	13	169
7	9	81
8	10	100
9	7	49
10	9	81
$n=10$	$\Sigma x=94$	$\Sigma x^2=934$

Средняя плодовитость в этой выборке равна:

$$M = \frac{\Sigma x}{n} = \frac{94}{10} = 9{,}4 \text{ поросенка.}$$

Среднее квадратическое отклонение вычисляется по формуле:

$$\sigma = \sqrt{\frac{\alpha}{n-1}}, \text{ где } \alpha = \Sigma x^2 - \frac{(\Sigma x)^2}{n}.$$

Подставим в эти формулы соответствущие значения:

$$\alpha = 934 - \frac{(94)^2}{10} = 934 - \frac{8836}{10} = 50{,}4;$$

$$\sigma = \sqrt{\frac{50{,}4}{10-1}} = \sqrt{\frac{50{,}4}{9}} = \sqrt{5{,}6} = 2{,}37 \text{ поросенка.}$$

5. ВЫЧИСЛЕНИЕ σ ДЛЯ МАЛЫХ ВЫБОРОК ПРИ ДРОБНЫХ ИЛИ МНОГОЗНАЧНЫХ ЗНАЧЕНИЯХ ВАРИАНТОВ

При вычислении σ при дробных или многозначных значениях варианта пользование приемом возведения вариантов в квадрат создает большие трудности, так как получаются громоздкие многозначные числа.

Поэтому для упрощения работы производят вычитание из каждого варианта условной средней A, близкой к средней арифметической. Такую условную среднюю выбирают грубо «на глаз». Полученные отклонения $(x-A)$ возводят в квадрат.

Формула среднего квадратического отклонения при этом выражается так:

$$\sigma = \sqrt{\frac{\alpha}{n-1}}, \text{ где } \alpha = \Sigma\Delta^2 - \frac{(\Sigma\Delta)^2}{n},$$

а формула средней арифметической: $M = A + \frac{\Sigma\Delta}{n}$.

Значение $\Delta = (x-A)$ обозначает величину отклонения варианта от условной средней A.

Пример. Определить изменчивость средней урожайности кустов картофеля на опытной делянке:

№ куста	Вес клубней с куста в *кг*, x	$(x-A)=\Delta$	Δ^2
1	5,2	−2,3	5,29
2	6,1	−1,4	1,96
3	10,0	+2,5	6,25
4	9,5	+2,0	4,00
5	8,7	+1,2	1,44
6	5,0	−2,5	6,25
7	7,0	−0,5	0,25
8	9,2	+1,7	2,89
9	8,0	+0,5	0,25
10	7,5	0	0
	$\Sigma x = 76{,}2$	$\Sigma\Delta = +1{,}2$	$\Sigma\Delta^2 = 28{,}58$

В качестве условной средней A берем урожайность, равную 7,5 *кг*. Вычитаем из каждого варианта x значение A: $5{,}2-7{,}5=-2{,}3$; $6{,}1-7{,}5=-1{,}4$ и т. д. Алгебраическая сумма $\Sigma(x-A)=\Sigma\Delta$ оказалась равна $+1{,}2$ *кг*. После этого составляем ряд квадратов отклонений (Δ^2).

Имея величины $\Sigma\Delta=+1{,}2$ и $\Sigma\Delta^2=28{,}58$, можно вычислить M и σ. Подставляем в формулу средней арифметической значения A и $\Sigma\Delta$:

$$M = A + \frac{\Sigma\Delta}{n} = 7{,}5 + \frac{1{,}2}{10} = 7{,}62 \text{ кг};$$

$$\alpha = \Sigma\Delta^2 - \frac{(\Sigma\Delta)^2}{n} = 28{,}58 - \frac{(1{,}2)^2}{10} =$$

$$= 28{,}58 - \frac{1{,}44}{10} = 28{,}58 - 0{,}144 = 28{,}436.$$

Подставляем полученное число в формулу среднего квадратического отклонения:

$$\sigma = \sqrt{\frac{\alpha}{n-1}} = \sqrt{\frac{28{,}436}{10-1}} = \sqrt{\frac{28{,}436}{9}} = \sqrt{3{,}159} = 1{,}78 \text{ кг}.$$

Таким образом, для нашего ряда изменчивость выразилась средним квадратическим отклонением, равным 1,78 *кг*.

6. СПОСОБ ПОЛУЧЕНИЯ СУММАРНОЙ СРЕДНЕЙ АРИФМЕТИЧЕСКОЙ M И СРЕДНЕГО КВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ σ

Имея несколько вариационных рядов, можно получить обобщенные среднюю арифметическую и среднее квадратическое отклонение на основе их частных значений при отсутствии данных σ самих вариационных рядов.

Формула суммарной средней арифметической выглядит так:

$$M = \frac{\Sigma (M_i \cdot n_i)}{\Sigma n_i}.$$

Для получения суммарного среднего квадратического отклонения пользуются следующей формулой:

$$\sigma = \sqrt{\frac{\Sigma n_i \sigma_i^2 + \Sigma (M_i - M)^2 \cdot n_i}{\Sigma n_i}}.$$

Пример. Определить суммарную среднюю арифметическую M и среднее квадратическое отклонение σ по показателю урожайности пшеницы четырех хозяйств:

№ хозяйства	1	2	3	4
Общая площадь в *га*, n_i	100	80	150	200
Средняя урожайность в *ц*. M_i	20	25	15	18
Среднее квадратическое отклонение, σ_i	1	2	3	5
$n_i \cdot M_i$	2000	2000	2250	3600
σ_i^2	1	4	9	25
$(M_i - M)$	1,42	6,42	—3,58	—0,58
$(M_i - M)^2$	2,016	41,22	12,82	0,34
$(M_i - M)^2\ n_i$	201,6	3297,6	1923,0	68,0
$n_i \cdot \sigma_i^2$	100	320	1350	5000

Для получения суммарной средней следует умножить каждую частную M_i на частное число наблюдений n_i, и полученные произведения сложить, т. е. $20 \cdot 100 + 25 \cdot 80 + 15 \cdot 150 + 18 \cdot 200 = 9850$, отсюда суммарная средняя выразится

$$M = \frac{\Sigma (M_i \cdot n_i)}{\Sigma n_i} = \frac{9850}{530} = 18,58\ ц.$$

Для сигмы следует получить отклонения частных средних M_i от суммарной средней M. Это отклонение по каждому хозяйству равно: для первого. $M_i - M = 20 - 18{,}58 = 1{,}42$ *ц*; для второго: $25 - 18{,}58 = 6{,}42$ *ц*; для третьего: $15 - 18{,}58 = -3{,}58$ *ц*; для четвертого $18 - 18{,}58 = -0{,}58$ *ц*. Полученные отклонения возводятся в квадрат, умножаются на частное число наблюдений и суммируются, т. е. $\Sigma (M_i - M)^2 \cdot n_i$. Этот расчет дает следующие показатели для каждого хозяйства: $1{,}42^2 \cdot 100 = 201{,}6$; $6{,}42^2 \cdot 80 = 3297{,}6$; $(-3{,}58)^2 \cdot 150 = 1923{,}0$; $(-0{,}58)^2 \cdot 200 = 68{,}0$.

$$\Sigma (M_i - M)^2 \cdot n_i = 201{,}6 + 3297{,}6 + 1923{,}0 + 68{,}0 = 5490{,}2.$$

После этого, производится определение суммы произведений частного числа наблюдений на частную сигму в квадрате $(n_i \cdot \sigma_i^2)$.

Из такого расчета находим:

$$100 \cdot 1 = 100; \quad 80 \cdot 4 = 320; \quad 150 \cdot 9 = 1350; \quad 200 \cdot 25 = 5000,$$

откуда: $\Sigma n_i \sigma_i^2 = 6770$.

Подставляем полученные значения в формулу суммарного среднего квадратического отклонения:

$$\sigma = \sqrt{\frac{\Sigma n_i \sigma_i^2 + \Sigma (M_i - M)^2 \cdot n_i}{\Sigma n_i}} = \sqrt{\frac{6770 + 5490{,}2}{530}} =$$

$$= \sqrt{23{,}13} = 4{,}806 \text{ ц}.$$

Таким образом, суммарные M и σ найдены по показателям их частных значений.

7. ВЫЧИСЛЕНИЕ СРЕДНЕГО КВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ σ ДЛЯ АЛЬТЕРНАТИВНЫХ ПРИЗНАКОВ

Вариационная статистика позволяет обрабатывать не только данные о количественных, но и о качественных признаках, т. е. таких, как цвет, форма, окраска и т. п. При этом часто качественные признаки могут быть выражены в виде альтернативных признаков, т. е. иметь только два состояния. Например: пол мужской или женский, остистые или безостые колосья злаков, пигментированный или альбиносный тип шерсти и т. п.

Для измерения изменчивости качественных признаков вычисление среднего квадратического отклонения σ производится по следующей формуле:

$$\sigma = \frac{\sqrt{p_1 \cdot p_2}}{n},$$

где p_1 — абсолютные показатели частот присутствия признака;

p_2 — абсолютные показатели частот отсутствия этого признака;

n — число наблюдений.

Пример. Определить изменчивость у помесного потомства кур по форме гребня:

Разновидность кур	Куры с розовидным гребнем, p_1	Куры с листовидным гребнем, p_2
Число голов	150	50
Выражение количества кур в % от общего числа голов	75	25

$$\sigma = \frac{\sqrt{p_1 \cdot p_2}}{n} = \frac{\sqrt{150 \cdot 50}}{200} = \frac{\sqrt{7500}}{200} = \frac{86,6}{200} \approx 0,43.$$

Умножая полученную дробь на 100, выражаем изменчивость в процентах: $0,43 \cdot 100 = 43\%$.

Этот же результат можно получить и по другой формуле, если имеется процентное соотношение частот:

$$\sigma = \sqrt{p_1(\%) \cdot p_2(\%)} = \sqrt{75 \cdot 25} = \sqrt{1875} \approx 43\%.$$

Обе приведенные формулы могут быть использованы и для тех случаев, когда имеется не два качественных признака, а несколько.

Следует иметь в виду, что для альтернативных признаков самое большое значение показателя изменчивости может быть равно 50%, т. е когда $p_1 = p_2 = n : 2$.

Например: если $p_1 = 100\%$, $p_2 = 0$, тогда $\sigma = \sqrt{100 \cdot 0} = 0$; если $p_1 = 10\%$, а $p_2 = 90\%$, тогда $\sigma = \sqrt{10 \cdot 90} = 30\%$; если $p_1 = 50\%$ и $p_2 = 50\%$, то $\sigma = \sqrt{50 \cdot 50} = 50\%$, т. е. показатель изменчивости имеет наибольшее возможное значение.

8. ИЗМЕРЕНИЕ ВАРИАЦИИ ИЛИ ИЗМЕНЧИВОСТИ С ПОМОЩЬЮ КОЭФФИЦИЕНТА ВАРИАЦИИ C (или V)

Ранее указывалось, что среднее квадратическое отклонение σ выражается в именованных величинах. Поэтому, если требуется сравнить степень изменчивости признаков разноименных (например, изменчивость веса с изменчивостью роста, плодовитость со скороспелостью и т. п.), то пользоваться именованной величиной σ нельзя, т. к. сравниваемые σ имеют разное наименование (*кг* и *см*, число голов и *кг* и т. п). В таких случаях вычисляют коэффициент вариации или изменчивости, в котором величина сигмы отнесена к средней арифметической и выражена в процентах:

$$C = \frac{\sigma}{M} \cdot 100.$$

В этом случае изменчивость выражается относительным показателем (%) и дает возможность сравнивать ее у разноименных вариантов. Коэффициент вариации показывает, насколько велика вариабильность по сравнению со средней арифметической.

Пример. Сравнить изменчивость живого веса и промера обхвата пясти у коров джерсейской породы.

Показатели по живому весу в *кг*

M_1 . . . . 390;

σ_1 . . . . 35.

$$C_1 = \frac{35}{390} \cdot 100 = 9\%.$$

Показатели по обхвату пясти в *см*:

M_2 . . . . 13;

σ_2 . . . . 0,05.

$$C_2 = \frac{0,05}{13} \cdot 100 = 0,4\%.$$

Изменчивость живого веса в 22 раза выше изменчивости такого скелетного промера, как обхват пясти.

Коэффициент изменчивости C для альтернативных признаков не вычисляется, а изменчивость их выражается в % через σ по указанной на стр. 48 формуле.

Пример. Определить изменчивость окраски плодов помидор во втором гибридном поколении.

Всего было 558 плодов, из них: 39 шт. ($\approx 7\%$) — желтых, 121 шт. ($\approx 22\%$) — розовых, 105 шт. ($\approx 19\%$) — красных, 293 шт. ($\approx 52\%$) белых.

Количество плодов, оставшееся после вычитания количества плодов данного сорта из общего числа плодов, будет составлять число плодов прочих окрасок по отношению к данной окраске и выразится следующими числами: 519 шт. (93%) по отношению к желтым, 437 шт. (78%) по отношению *к розовым, 453 шт. (81%) по отношению к красным, 265 шт.* (48%) по отношению к белым.

После этого можно определить изменчивость окраски плодов помидор:

$$\sigma = \sqrt{7 \cdot 93} \approx 25\% \text{ (для желтых плодов)},$$

$$\sigma = \sqrt{22 \cdot 78} \approx 41\% \text{ (для розовых плодов)},$$

$$\sigma = \sqrt{19 \cdot 81} \approx 39\% \text{ (для красных плодов)},$$

$$\sigma = \sqrt{52 \cdot 48} \approx 50\% \text{ (для белых плодов)}.$$

Наибольшая изменчивость оказалась по белоплодной форме, где $\sigma \approx 50\%$, наименьшая изменчивость — по желтоплодной форме, где $\sigma \approx 25\%$.

ГЛАВА V

ГРАФИЧЕСКИЙ МЕТОД АНАЛИЗА ВАРИАЦИОННЫХ РЯДОВ И ТИПЫ РАСПРЕДЕЛЕНИЙ (нормальное, биноминальное, Пуассона, Шарлье и трансгрессии)

Анализ статистических совокупностей, сформированных в вариационный ряд, может быть углублен путем использования графического метода в виде построения эмпирических и теоретических вариационных кривых.

1. ПОСТРОЕНИЕ ЭМПИРИЧЕСКОЙ ВАРИАЦИОННОЙ КРИВОЙ ПРИ ПРЕРЫВИСТОМ ЗНАЧЕНИИ ПРИЗНАКА

Если признак имеет прерывистый характер (число зерен, яиц, крольчаг и т. п., т. е. когда признак неделим), то графически кривая изображается так: по оси абсцисс откладываются отрезки, характеризующие величину признака (показатели вариантов x), по оси ординат — частоты p.

На рис. 1 рассмотрим плодовитость крольчих на один помет. Пусть плодовитость x и частоты p соответственно выразятся:

x	0	1	2	3	4	5	6	7
p	0	10	15	25	30	15	5	0

Полученная ломаная кривая называется полигоном распределения.

2. ПОСТРОЕНИЕ КРИВОЙ С НЕПРЕРЫВНЫМ ЗНАЧЕНИЕМ ПРИЗНАКА

Если признак имеет непрерывный характер (выражается в *см, кг, л*), то графическое изображение ряда ведется так: по оси абсцисс откладываются варианты классов x, по оси ординат — частоты p. Отрезки на оси x служат основанием прямоугольников, которые строятся в высоту, соответствующую значению частот p. Построенные таким способом прямоугольники образуют **гистограмму**, характеризующую вариационный ряд.

Пример. Построить гистограмму высоты растений по следующему вариационному ряду (рис. 2):

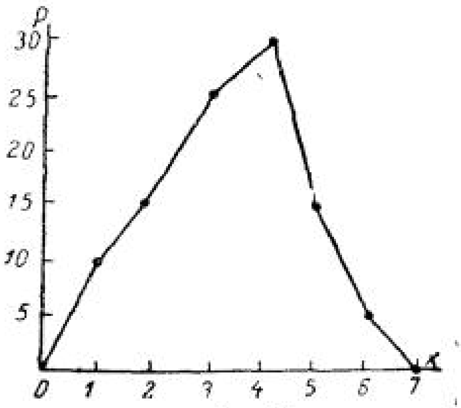

Рис. 1

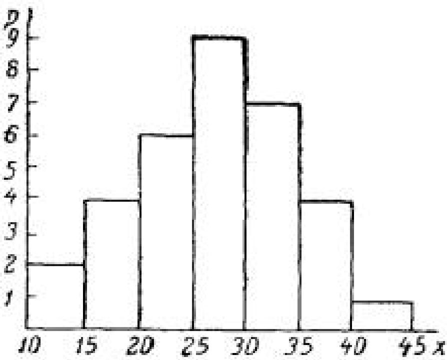

Рис. 2

x (высота в *см*)	10—15	15—20	20—25	25—30	30—35	35—40	40—45
p (частоты)	2	4	6	9	7	4	1

Каждый эмпирический вариационный ряд, построенный на ограниченном числе наблюдений, при бесконечно большом числе наблюдений превращается в теоретический ряд, частоты которого при графическом построении дают плавную вариационную кривую, отражающую закономерности данной совокупности.

Таким образом, каждому эмпирическому ряду соответствует теоретическая кривая распределения.

Многообразие эмпирических кривых и их теоретических кривых распределения может быть подразделено на различные типы. Каждый тип кривых отражает особую закономерность в распределении членов совокупности по вариантам, которых выделено уже значительное количество. Остановимся на таких типах кривых и рядах распределений, которые чаще всего встречаются при статистической обработке материалов, а именно: нормальном и биноминальном распределениях, распределениях Пуассона и Шарлье и трансгрессивных рядах распределения.

3. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ И НОРМАЛЬНАЯ КРИВАЯ

Если случайная переменная x (варьирующий признак) может принимать значения от $-\infty$ до $+\infty$ и если вероятность ее появления выражается формулой:

$$P_x = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}},$$

то это означает, что переменная x имеет нормальное распределение.

В этой формуле P_x —вероятность появления переменной x; σ — среднее квадратическое отклонение; π — постоянное число, равное 3,14159; e — основание натуральных логарифмов, равное 2,71828; x — варьирующий признак.

Нормальная кривая распределения изображена на рис. 3.

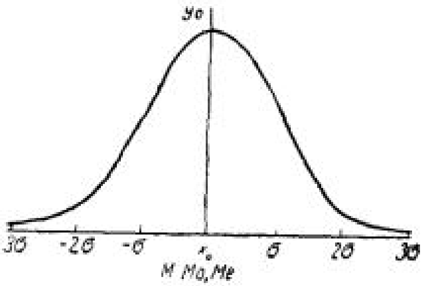

Рис. 3

То значение x, которое соответствует величине средней арифметической, берется в качестве начальной точки x_0; вправо от нее откладываются значения x, превышающие величину средней арифметической и доходящие до максимума ($x_{макс}$), а влево откладываются значения x, меньшие величины средней арифметической и доходящие до минимума ($x_{мин}$).

Ось ординат служит для отложения значений частот, которые при $n \to \infty$ превращаются в значения вероятностей P и равны отрезкам перпендикуляров, восставленных из соответствующих значений x. Перпендикуляр, восставленный из x_0, соответствует вершине кривой, т. е. максимальному значению ординаты и обозначается через y_0.

Если соединить концы отрезков перпендикуляров линией, то она и будет служить нормальной кривой. Форма нормальной кривой колоколообразная, плавная.

Уже при первоначальном рассмотрении полученной кривой заметен ряд ее закономерностей. Кривая имеет симметричное расположение вероятностей по отношению к нулевой ординате y_0. Это означает, что точка x_0 совмещает в нормальной кривой значение варианта, равного значениям средней арифметической, моды и медианы, т. е. $x_0 = M = Mo = Me$.

Для кривой характерно, что чем дальше отстоит значение вариантов от значения средней арифметической, тем меньше их вероятности. Постепенно в бесконечности ветви кривой плавно и асимптотически приближаются к оси абсцисс.

Основанием площади, ограниченной кривой, служит ось абсцисс, на которой нанесены значения x, отклоняющиеся от средней арифметической вправо и влево. Размах варьирования от минимального до максимального значений варианта в нормальном распределении почти полностью ограничен интервалом $\pm 3\sigma$ от средней арифметической (что видно из рис. 3). Площадь, ограниченная нормальной кривой, принимается за единицу. Площадь, ограниченная ординатами, соответствующими отклонениям x на $+3\sigma$ и -3σ, составляет 0,997 от всей площади, ограниченной нормальной кривой. Это означает, что если площадь, ограниченную всей кривой и осью x, принять за 100%, то на интервал от $+3\sigma$ до -3σ приходится 99,7% всей площади (или 99,7% числа наблюдений всей совокупности).

Исходя из этой функции, отвечающей нормальной кривой, можно определить, какая доля (или %) площади, а следовательно, и какой процент (или какое количество) всех наблюдений находятся в пределах, ограниченных теми или иными ординатами нормальной кривой на том или ином расстоянии по оси абсцисс от точки x_0.

На рис. 4 изображены схемы задач на нахождение доли частот той или иной части кривой. Заштрихованная часть означает искомую площадь. Так, на рис. *4а* требуется найти процент частот в границах от x_0 до $+3\sigma$; на рис. *4б* требуется найти процент частот, расположенных в границах от x_0 до $-1{,}5\sigma$; на рис. *4в* отыскивается процент частот в границах от $+2\sigma$ до $+3\sigma$; на рис. *4г* определяется процент частот от -2σ до -3σ; на рис. *4д* определяется процент частот в границах от -2σ до $+2\sigma$ и на рис. *4е* требуется определить частоты, находящиеся за пределами $-\sigma$ и $+\sigma$.

Для того, чтобы найти площадь, ограниченную кривой, осью абсцисс и ординатами, следует пользоваться таблицей интеграла вероятностей (стр. 56, табл. 4). С помощью интегрирования уравнения нормальной кривой можно определить те или иные площади, ограниченные кривой, ординатами и осью x при решении ряда задач и в частности для тех случаев, которые схематично приведены на рис. 4 *а, б, в, г, д, е*.

Таблица интеграла вероятностей состоит из двух граф. Одна из них дает показатели отклонения варианта от сред-

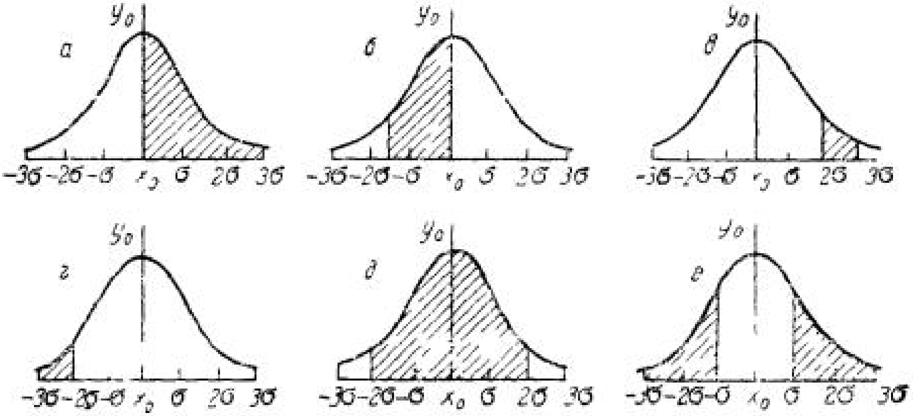

Рис. 4

ней арифметической, выраженного в долях σ; вторая графа выражает площадь, ограниченную кривой и ординатами при заданном значении отклоненчя x от $\bar{x}$

Если, например, требуется узнать, какой процент от общего числа наблюдений будут составлять объекты, отклоняющиеся от средней арифметической на $-1{,}5\sigma$, то следует найти в таблице интеграла вероятностей строчку со значением $\sigma=1{,}5$ и посмотреть, какое значение площади соответствует этому отклонению Оказывается, что для $(x-\bar{x})$, равного $1{,}5\sigma$, площадь, ограниченная кривой и ординатами, равна 0,43319 Это означает, что 43,319% всех особей данной совокупности будут иметь величину варьирующего признака в пределах от $\bar{x}$ до $\bar{x}-1{,}5\sigma$ (рис. 4 *б*).

Пример. Определить численность помесных коров в данном районе, у которых жирномолочность превышает средний уровень на $+2\sigma$ При этом известно, что число помесных коров в районе равно 1000 голов, их средняя жирномолочность равна 4,0% жира в молоке и среднее квадратическое отклонение этого показателя равно 0,5% Этот пример соответствует схеме, изображенной на рис 4*в*.

По таблице интеграла вероятностей оказывается, что отклонению x от средней на $+2\sigma$ соответствует площадь, равная 0,47725. Это означает, что 47,725% коров имеют жирномолочность в пределах $\bar{x}+x$, равных от 4,0 до 5,0%, т. е. в границах от $\bar{x}$ до 2σ. Процент же коров с жирномолочностью свыше $+2\sigma$ получается из графика путем вычитания из площади,

Таблица 4

Интеграл вероятностей (площадь ограниченная нормальной кривой и ординатами y_0 и y_x при отклонении вариантов x в долях σ от $\bar{x}$)[1]

Отклонение в долях σ	Площадь между максимальной ординатой и ординатой при $x\sigma$	Отклонение в долях σ	Площадь между максимальной ординатой и ординатой при $x\sigma$	Отклонение в долях σ	Площадь между максимальной ординатой и ординатой при $x\sigma$	Отклонение в долях σ	Площадь между максимальной ординатой и ординатой при $x\sigma$
0,0	0,00000	1,1	0,36433	2,2	0,48610	3,3	0,49952
0,1	0,03983	1,2	0,38493	2,3	0,48928	3,4	0,49966
0,2	0,07926	1,3	0,40320	2,4	0,49180	3,5	0,49977
0,3	0,11791	1,4	0,41924	2,5	0,49379	3,6	0,49984
0,4	0,15542	1,5	0,43319	2,6	0,49534	3,7	0,49989
0,5	0,19146	1,6	0,44520	2,7	0,49653	3,8	0,49993
0,6	0,22575	1,7	0,45543	2,8	0,49744	3,9	0,49995
0,7	0,25804	1,8	0,46407	2,9	0,49813	4,0	0,49997
0,8	0,28814	1,9	0,47128	3,0	0,49865		
0,9	0,31594	2,0	0,47725	3,1	0,49903		
1,0	0,34134	2,1	0,48214	3,2	0,49931		

равной 50% всей площади, процента площади с отклонением признака x на $+2\sigma$, т. е. 47,725%. Следовательно, 50,0—47,725=2,275%. Это означает, что примерно 23 коровы из 1000 будут иметь жирномолочность, превышающую $+2\sigma$. Так как $\bar{x}=4,0\%$, а $2\sigma=0,5\cdot2=1,0\%$, то группа помесных коров за пределами $+2\sigma$ будет иметь процент жира в молоке от 5% и выше.

Данные о площадях, ограниченных нормальной кривой, осью абсцисс и ординатами, в таблице интеграла вероятностей будут так же служить показателями вероятности появления варианта с тем или иным отклонением в долях σ от средней арифметической $\bar{x}$.

Из последнего разобранного примера следует, что вероятность появления помесных коров, имеющих жирномо-

[1] Таблица взята в сокращенном виде с интервалами в 0,1 σ из книги Ф. Миллса. «Статистические методы», Госстатиздат, М., 1958, стр. 769.

лочность свыше 5%, соответствующую отклонению от $\bar{x}$ на $+2\sigma$, составляет $P = 0{,}0228$ (т. к. $0{,}50-0{,}4772 = 0{,}0228$).

Приведенный пример показывает, что закономерности нормальной кривой, выраженные через интеграл вероятностей, могут быть использованы для различных расчетов.

Пользуясь уравнением $P_x = y_x = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}}$, можно определить значения теоритических частот любого вариационного ряда и сопоставить их с эмпирическими частотами.

Выше было уже отмечено, что *площадь, образованная* кривой и осью абсцисс, равна единице. Но при вычислении ординат для конкретного вариационного ряда в числитель формулы подставляется величина, соответствующая числу наблюдений данного ряда, и тогда формула принимает такой вид:

$$y_x = \frac{n}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}}.$$

В этой формуле выражение $\frac{n}{\sigma\sqrt{2\pi}}$ соответствует значению ординаты y_0 в точке x_0, т. е. $\bar{x}$. Определив по формуле величину y_0, можно воспользоваться следующим преобразованным уравнением нормальной кривой: $y_x = y_0 \cdot e^{-\frac{x^2}{2\sigma^2}}$. Отсюда значение любой ординаты данного вариационного ряда соответствующее частотам p при всех значениях x, может быть получено путем умножения величины максимальной ординаты на выражение $e^{-\frac{x^2}{2\sigma^2}}$.

Для упрощения расчетной работы можно пользоваться подсобной табл. 5, показывающей значение ординаты при определенном значении x. Ординаты y в нормальной кривой распределения являются функцией x и могут быть заменены в формуле $y = \frac{n}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}}$ несколько иным выражением, если принять, что $\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}} = f(x)$.

Величина $f(x)$ при определенном значении x, выраженном в отклонениях σ, может быть найдена по табл. 5.

Таким образом, рабочая формула для определения ординат нормальной кривой, соответствующих теоретическим частотам вариационного ряда, будет выглядеть следующим образом:

$$y_x = \frac{n}{\sigma} \cdot f(x).$$

Разберем пример вычисления теоретических частот.

Таблица 5[5]

Функция нормированного отклонения $\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}}$.

Отклонение x от $\overline{x}$ (в долях σ)	$f(x)$	Отклонение x от $\overline{x}$ (в долях σ)	$f(x)$	Отклонение x от $\overline{x}$ (в долях σ)	$f(x)$
0,00	0,39894	1,40	0,14973	2,80	0,00792
0,10	0,39695	1,50	0,12952	2,90	0,00595
0,20	0,39104	1,60	0,11092	3,00	0,00443
0,30	0,38139	1,70	0,09405	3,10	0,00327
0,40	0,36827	1,80	0,07895	3,20	0,00238
0,50	0,35207	1,90	0,06562	3,30	0,00172
0,60	0,33322	2,00	0,05399	3,40	0,00123
0,70	0,31225	2,10	0,04398	3,50	0,00087
0,80	0,28969	2,20	0,03547	3,60	0,00061
0,90	0,26609	2,30	0,02833	3,70	0,00042
1,00	0,24197	2,40	0,02239	3,80	0,00029
1,10	0,21785	2,50	0,01753	3,90	0,00020
1,20	0,19419	2,60	0,01358	4,00	0,00014
1,30	0,17137	2,70	0,01042		

Пример. Имеется вариационный ряд по показателю жирномолочности коров джерсейской породы (табл. 6). Определить с помощью таблицы значений ординат нормальной кривой теоретические частоты этого ряда. Известно, что $n = 150$; $M = 5{,}792\%$; σ^* (выражена в классовом промежутке) равна 2,12; $\sigma = 2{,}12 \cdot 0{,}2 = 0{,}424\%$; $\frac{n}{\sigma^*} = 150 : 2{,}12 = 70{,}75$.

Для того, чтобы найти теоретические частоты данного ряда, следует найти нормированное отклонение варианта x от средней арифметической $(x - M)$, и полученное значение разделить на среднее квадратическое отклонение σ. Эти расчеты

[5] А. К. Митропольский. Техника статистического исчисления. Сельколхозгиз, М. — Л., 1931.

приведены в 3 и 4 графах табл. 6. В графу 5 вносятся данные значений $f(x)$ из таблицы значений ординат нормального распределения (табл. 5, стр. 58). После этого следует найти значение $\frac{n}{\sigma^*}$, которое входит в формулу нормального распределения. При этом значение сигмы берется *не в именован-*

Таблица 6

Обработка эмпирического вариационного ряда для вычисления теоретических частот

x (середина класса)	Эмпирические частоты	$(x - M) =$ $= (x - 5{,}792)$	$\frac{x-M}{\sigma} =$ $= \frac{x - 5{,}792}{0{,}421}$	$f(x)$ по таблице	$\frac{n}{\sigma^*} \cdot f_x =$ $= 70{,}75 \cdot f_x$	Теоретические частоты (после округления)
1	2	3	4	5	6	7
4,6	1	−1,192	−2,811	0,00770	0,545	1
4,8	1	−0,992	−2,339	0,02582	1,827	2
5,0	5	−0,792	−1,868	0,06943	4,912	5
5,2	12	−0,592	−1,396	0,14937	10,568	11
5,4	15	−0,392	−0,925	0,26129	18,486	18
5,6	15	−0,192	−0,453	0,36058	25,511	25
5,8	50	+0,008	+0,019	0,39894	28,225	28
6,0	20	+0,208	+0,491	0,35381	25,032	25
6,2	15	+0,408	+0,962	0,25164	17,803	18
6,4	10	+0,608	+1,433	0,14350	10,153	10
6,6	4	+0,808	+1,908	0,06438	4,558	5
6,8	2	+1,008	+2,373	0,02349	1,662	2
	$n = 150$					$n = 150$

ных величинах, а в относительных (σ^*), т. е. без умножения *на величину класса к. В нашем примере относительное значение* сигмы равно 2,12. Разделив n на σ^*, получаем 70,75. В графе 6 производится умножение этой величины на значение $f(x)$, т .е. $\frac{n}{\sigma^*} \cdot f(x)$. Полученные показатели и будут являться теоретическими частотами. Производим их округление до целых и получаем окончательные данные ряда теоретических частот.

Между эмпирическим и теоретическим рядами имеются некоторые различия. Если изобразить эмпирическую и теоретическую кривые по данным, полученным после расчетов, приведенных в табл. 6, то оказывается, что у эмпирического ряда имеет место вытянутая вершина (так называемый эксцесс), у теоретического ряда этого не наблюдается (рис. 5).

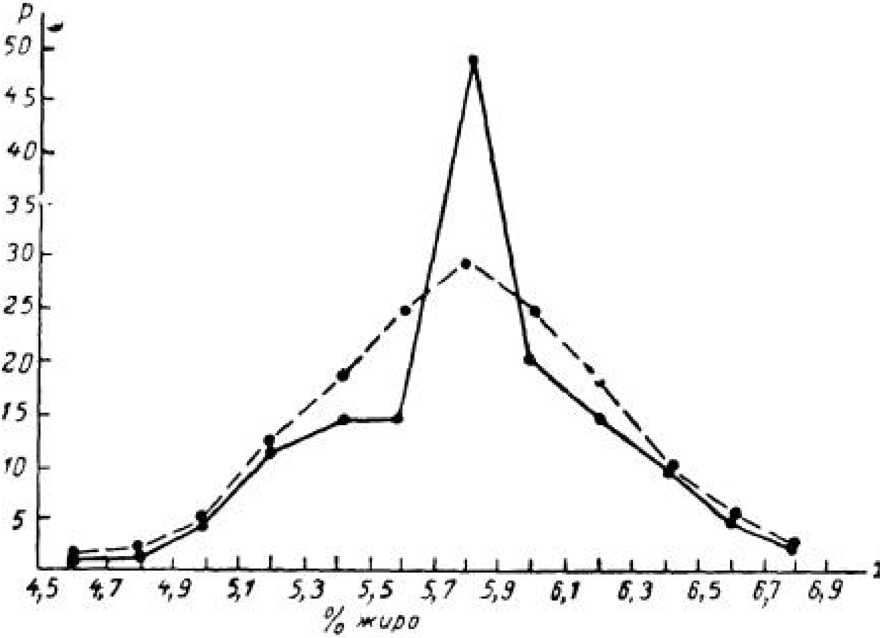

Рис. 5

4. БИНОМИНАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Биноминальное распределение имеет место в тех случаях, когда при проведении n опытов вероятность появления события A равна P, а вероятность его непоявления равна $1-P=Q$. Теоретические частоты появления события A $(0, 1, 2, \ldots, m$ раз) равны членам разложения бинома Ньютона $(P+Q)^m$, где m — число наблюдений в каждом опыте.

Такая особенность биноминального распределения указывает на то, что оно будет иметь место при альтернативных признаках. В биологии нередко встречаются случаи, когда распределение качественных признаков имеет биноминальное распределение.

Биноминальное распределение прерывисто и поэтому на графике изображается ломаной линией. Форма графика зависит от значений P и m.

Биноминальное распределение и его график будут симметричны, если $P=Q$. При увеличении числа наблюдений m в опытах биноминальные кривые становятся симметричными.

Средней арифметической в биноминальном распределении служит произведение mP, т. е. произведение числа наблюдений на вероятность появления события. Дисперсия биноминального распределения $\sigma^2 = mP \cdot Q$ выражается как произведение числа повторений в опыте на вероятность появления и вероятность непоявления данного события. Отсюда среднее квадратическое отклонение $\sigma = \sqrt{mPQ}$.

5. РАСПРЕДЕЛЕНИЕ ПУАССОНА

Распределение Пуассона относится к числу появлений редких (маловероятных) событий в большом числе опытов.

Для этого распределения характерно, что варьирующий признак x может принимать только целые значения, такие, как 0, 1, 2, 3 и т. д.

В распределении Пуассона среднее значение признака x совпадает с величиной дисперсии σ^2; таким образом, для него характерно наличие одного параметра $\bar{x}$ (или M), в то время, как нормальное распределение имеет два параметра, полностью его характеризующих, — $\bar{x}$ (или M) и σ.

Поэтому, если в распределении, состоящем из целых чисел, средняя и дисперсия отличаются друг от друга незначительно, то такое распределение можно считать близким к типу распределения Пуассона.

Распределение Пуассона в биологии используется для многих случаев и, прежде всего, для редко наступающих событий, таких как появление в потомстве уродов, рождение троень у одноплодных животных, образование колоний микроорганизмов в чашке Петри и т. п.[6].

6. АСИММЕТРИЧНЫЕ ВАРИАЦИОННЫЕ КРИВЫЕ И РАСПРЕДЕЛЕНИЕ ШАРЛЬЕ. МЕРА АСИММЕТРИИ As

В отличие от симметричных нормальных кривых встречаются такие распределения, у которых по одну сторону от модального варианта частоты уменьшаются более быстро, чем по другую и происходит перемещение вершины кривой в пра-

[6] В. И. Романовский. Применения математической статистики в опытном деле. М. — Л., 1947, стр. 68—69.

вую или левую сторону от средней арифметической — это создает заметную асимметрию вариационного ряда или, как говорят, ее косость. Такие кривые получили название асимметричных кривых, или кривых распределения Шарлье. Их общий вид выглядит следующим образом (рис. 6 и 7):

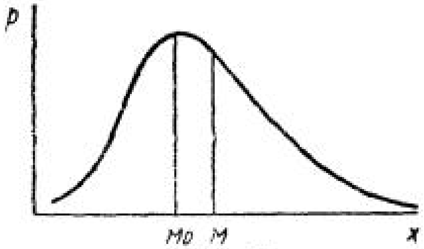

Рис. 6

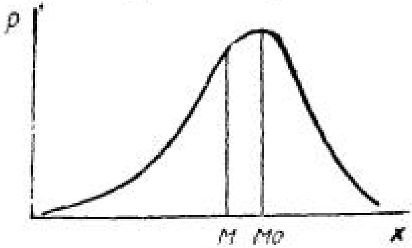

Рис. 7

Асимметрия может быть положительной, или правосторонней, когда M лежит правее моды Mo (рис. 6), и отрицательной, или левосторонней, когда M лежит левее Mo (рис. 7).

В общей форме асимметрия выражается так:

$$As = \frac{M - Mo}{\sigma} \text{ или } As = \frac{3\,(M - Me)}{\sigma}.$$

Это означает отношение разности между значениями средней арифметической и величиной моды или медианы к средним квадратическим отклонениям. При этом (приблизительно)

$$Me = \frac{Mo + 2M}{3} \text{ и } M = \frac{3Me - Mo}{2}.$$

Для вычисления точного значения степени асимметрии пользуются третьим центральным моментом μ_3, при использовании которого формула выглядит следующим образом:

$$As = \frac{\mu_3}{\sigma^3}.$$

Для этих формул σ берется в относительных величинах, т. е. без умножения на классовый промежуток κ.

Причинами, вызывающими асимметричное распределение, могут служить:

а) неправильно проведенная выборка, когда в нее вошло непропорционально много (или мало) представителей варианта с бóльшим или мéньшим их значением. Такие материалы представляют брак и к обработке не пригодны;

б) воздействия определенных факторов, сдвигающих частоту варьирующего признака в ту или другую сторону от среднего значения. Так, например, в генетических работах часто ставится цель опыта — получить сдвиг в распределении и вызвать положительную асимметрию, сопровождающую накопление частот у более высоких значений вариантов. Поэтому в таких случаях особенно важно выявить степень асимметрии и доказать, что это не случайное явление, вызванное методическими оплошностями при составлении выборки.

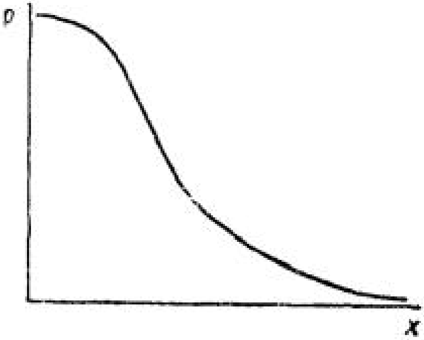

Рис. 8

Считается, что кривые при значении коэффициента асимметрии от 0,25 до 0,5 имеют умеренную косость.

При $As > 0$ — асимметрия положительная; при $As < 0$ — асимметрия отрицательная; у симметричных рядов $As = 0$.

Наиболее резко асимметрия проявляется в односторонних, или половинных, кривых (рис. 8).

С таким явлением встретился де Фриз при описании числа лепестков у калужницы.

7. ЭКСЦЕССИВНЫЕ ВАРИАЦИОННЫЕ КРИВЫЕ. МЕРА ЭКСЦЕССА *Ex*

Эксцессивные, или высоковершинные, кривые имеют особые свойства, отличающие их от кривых другого типа. Их общий вид изображен на рис. 9. Из рисунка видно, что эксцесс характеризуется скоплением большого количества частот около тех значений варьирующего признака, которые близки к значениям средней арифметической, моды и медианы. Модальный вариант при этом имеет значительно бо́льшее число наблюдений, чем в нормальных кривых распределения. В результате указанного эксцессивная кривая имеет сильно вытянутую узкую вершинную часть и удлиненные пологие боковые ветви.

Эта особенность вариационных кривых была впервые описана ботаником Людвигом при исследовании некоторых растений.

Особенность эксцессивных кривых состоит еще и в том, что их крайние варианты ($x_{макс}$ и $x_{мин}$) отстоят от M не на

$\pm 3\sigma$. За пределами $\pm 3\sigma$ в этих кривых находится не 0,3% наблюдений, как это имело место в нормальных кривых, а 2—3% от числа наблюдений.

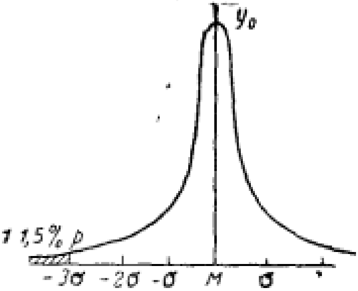

Рис. 9

Формула, позволяющая определить степень эксцесса (меру крутости), такова:

$$Ex = \frac{\mu_4}{\sigma^4} - 3, \text{ где}$$

$$\mu_4 = \frac{\Sigma pa^4}{n} - 4\frac{\Sigma pa^3}{n}\cdot\frac{\Sigma pa}{n} + $$

$$+ 6\frac{\Sigma pa^2}{n}\left(\frac{\Sigma pa}{n}\right)^2 - 3\left(\frac{\Sigma pa}{n}\right)^4$$

Для этой формулы среднее квадратическое отклонение берется в относительных величинах, т. е. без умножения на κ.

Ю. А. Филипченко приводит такой эксцессивный ряд в отношении числа каналов у гидромедузы:

Число каналов	2	3	4	5	6	7	8
Число особей, p	1	8	56	860	64	6	1
a	—3	—2	—1	0	+1	+2	+3
pa	—3	—16	—56	0	64	12	3
pa^2	9	32	56	0	64	24	9
pa^3	—27	—64	—56	0	64	48	27
pa^4	81	128	56	0	64	96	81

$$M = 5{,}004;\ n = 996;\ \mu_4 = \frac{\Sigma pa^4}{n} = \frac{506}{996} = 0{,}508;\ \sigma = 0{,}441.$$

(Для упрощения расчетов μ_4 берется по первой дроби, так как остальными можно пренебречь).

$$Ex = \frac{\mu_4}{\sigma^4} - 3 = \frac{0{,}508}{(0{,}441)^4} - 3 = 13{,}431 - 3 = 10{,}431.$$

Условно считают наличие незначительного эксцесса, если Ex меньше +0,4.

Если же частоты расположены в более узких пределах (т. е. в пределах меньше $+3\sigma$ и больше -3σ), то такой тип вариационного ряда называется дефектом, кривая в этом случае имеет плоскую вершину по сравнению с вершиной нормального распределения (рис. 10).

Эксцесс может доходить до нуля и принимать значения меньше нуля — отрицательный эксцесс (рис. 11). В этом случае форма кривой делается плосковершинной или двухвершинной

Значение отрицательного эксцесса не может быть меньше —2, так как при —2 получаются уже две самостоятельные кривые

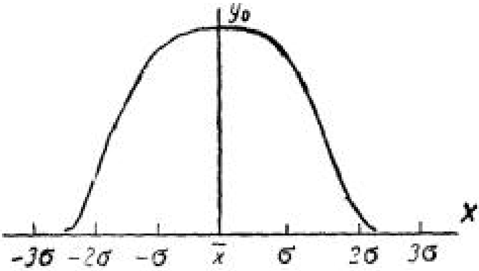

Рис 10

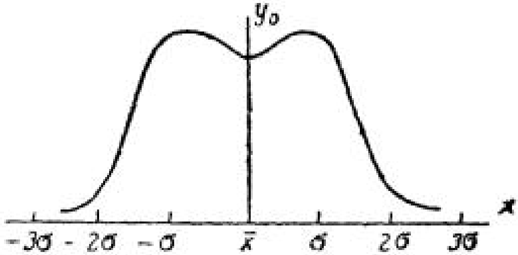

Рис 11

Так как меры эксцесса и асимметрии являются величинами относительными, а не именованными, величина классового промежутка k не влияет на их величину.

Причины эксцесса те же, что и для асимметрии, т е. неправильно сделанная выборка или воздействие какого-либо фактора на организмы изучаемой группы, вызывающее специфическое изменение частот в вариационном ряду

8. ДВУХВЕРШИННЫЕ И МНОГОВЕРШИННЫЕ КРИВЫЕ

Двухвершинные и многовершинные кривые свидетельствуют о том, что совокупность состоит не из однородного, а из разнородного материала Причиной этого может быть появление в обследуемой популяции объектов, у которых формируется новое свойство (видовое, породное, расовое и т. п). Поэтому в экспериментальной работе, когда основная задача исследования сводится к получению новых форм или отдельных признаков под воздействием исследуемого фактора (облучение, опыление смесью пыльцы, вегетативная гибридизация и т п), появление двухвершинности будет указывать на результативность применяемого фактора

Причину двухвершинности можно выяснить только применяя биологический метод анализа экспериментального материала.

9. ТРАНСГРЕССИВНАЯ ИЗМЕНЧИВОСТЬ

Под трансгрессивными вариационными рядами подразумеваются ряды, которые отличаются друг от друга своими средними арифметическими и крайний максимальный вариант одного из которых заходит за границы минимального варианта другого.

Это явление было описано несколькими биологами на различных признаках, в частности, Давенпортом и Бленкиншином на ругозе, де Фризом — на энотере, Гейнеке — на сельдях и т. д.

Пример. Из работ де Фриза по изучению длины лепестков у энотеры бьеннис и энотеры мурикатa (рис. 12) известно:

Число лепестков x	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Enotera murica-ta, p	1	3	8	12	14	16	3	—	—	—	—	—	—	—	—	—
Enotera biennis, p	—	—	—	—	1	4	6	6	9	9	12	5	3	1	6	1

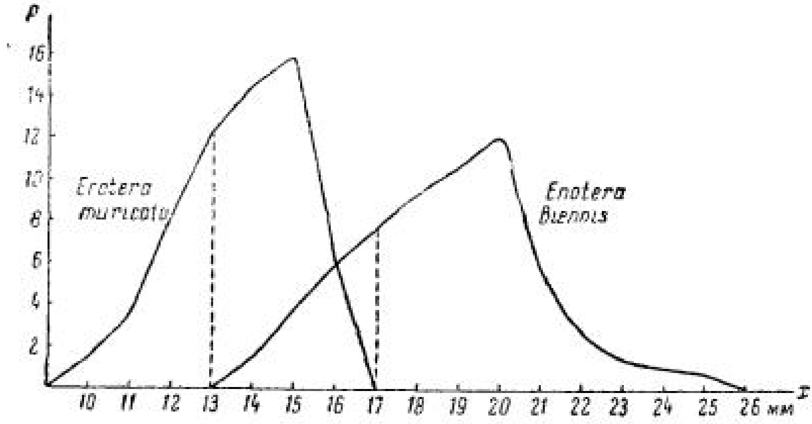

Рис. 12

Из рис. 12 видно, что растения, имеющие длину лепестков от 13 до 17 *мм*, могут быть отнесены как к вариационному ряду энотеры мурикатa, так и к ряду энотеры бьеннис.

Трансгрессивные ряды могут дать типичную двухвершинную кривую, но иногда смещение двух трансгрессивных рядов может дать и не двухвершинную, а одновершинную кривую.

Всегда надо помнить, что двухвершинность может иметь своей причиной не трансгрессивную изменчивость, а происходить под влиянием появления в совокупности особей с новыми качествами, вызванными разными факторами.

При наличии трансгрессивной изменчивости необходимо, следовательно, выяснить:

а) достоверна ли разница между средними арифметическими обоих рядов. Если разница достоверна, то это доказывает наличие трансгрессии, т. е. выборка состоит из неоднородного материала двух трансгрессирующих вариационных рядов. Об этом способе определения достоверности разницы между средними арифметическими будет сообщено позднее в разделе статистических ошибок;

б) определить принадлежность особей к тому или иному трансгрессивному ряду. Для этих целей ихтиологом Гейнеке при работе на сельдях предложен метод комбинированных признаков.

10. МЕТОД КОМБИНИРОВАННЫХ ПРИЗНАКОВ

Этот метод основывается на том, что сумма квадратов отклонений всех учитываемых признаков особей от средней величины каждого признака, свойственного данной расе, виду или группе, будет всегда величиной наименьшей. Таким образом, здесь используется свойство суммы квадратов отклонений $\Sigma(x-\tilde{x})^2 \rightarrow \min$.

Разберем это на примере: если известны средние показатели роста, веса, удоя, промеров тела и т. п. для нескольких пород или групп крупного рогатого скота, связанных трансгрессивной изменчивостью, то принадлежность каждого конкретного животного к той или другой породе может быть определена на основании того, будут ли показатели изучаемых признаков этого животного составлять минимальные значения суммы квадратов отклонений от средних величин одной из них. При этом необходимо учитывать, что для комбинированного метода следует брать некоррелированные друг с другом признаки.

Пример. Выяснить, к каким породам или помесям относится данная корова, если в отношении ее и в отношении коров исходных пород и помесных животных известны три показателя: живой вес, обхват груди и обхват пясти.

Средние арифметические по показателям были следующие:

Породность коров	Живой вес в *кг*	Обхват груди в *см*	Обхват пясти в *см*
Джерсейская	400	162	16
Остфризская	550	180	18
Помесная	480	170	17

Требуется определить, к какому трансгрессивному ряду отнести корову «*А*» неизвестного происхождения, имеющую следующие показатели: живой вес 490 *кг*, обхват груди 175 *см*, обхват пясти 18 *см*.

Для этого проводится сопоставление этих трех показателей данного животного с соответствующими средними показателями признаков каждой группы животных (джерсейской, остфризской и помесной), т. е. $(x_A - M)$ джерсеев, $(x_A - M)$ остфризов, $(x_A - M)$ помесей. Каждую из полученных разностей возводим в квадрат: $(x_A - M)^2$.

	Джерсеев	Остфризов	Помесей
	$(x_A - M)$	$(x_A - M)$	$(x_A - M)$
Вес в *кг* . . .	490—400=+90	490—550=—60	490—480=+10
Обхват груди в *см*	175—162=+13	175—180= —5	175—170=+5
Обхват пясти в *см*	18—16 =+ 2	18—18 = 0	18—17 =+1
	$(x_A - M)^2$	$(x_A - M)^2$	$(x_A - M)^2$
	$90^2 = 8100$	$(-60)^2=3600$	$10^2 = 100$
	$13^2 = 169$	$(-5)^2 = 25$	$5^2 = 25$
	$2^2 = 4$	$0^2 = 0$	$1^2 = 1$
	$\Sigma(x_A - M)^2 = 8273$	$\Sigma(x_A - M)^2 = 3625$	$\Sigma(x_A - M)^2 = 126$

Сумма квадратов отклонений показателей данного животного от средних показателей групп оказалась различной. Больше всего данное животное отличалось от показателей джерсейских коров, что видно из того, что $\Sigma(x_A - M)^2 = 8273$

была самой большой. Наименьшее отклонение было получено при сравнении данного животного с помесной группой, где $\Sigma(x_A - M)^2 = 126$.

Следовательно, корову «*A*» можно отнести к помесям.

Трансгрессивная изменчивость и метод комбинированных признаков особенно удобны для тех случаев, когда исследователь располагает данными о единичных экземплярах.

Такими единичными экземплярами, подвергаемыми изучению по комплексу признаков в генетических исследованиях, могут быть отдельные или малочисленные объекты, полученные от вегетативных прививок, отдаленного скрещивания и т. п.

Особый интерес этот метод представляет при работе с ископаемым материалом, когда зачастую мы располагаем только одним экземпляром или часто даже только некоторыми частями его для суждения о вымершем виде.

ГЛАВА VI

СТАТИСТИЧЕСКИЕ СВЯЗИ И МЕТОДЫ ИХ ВЫЧИСЛЕНИЯ

1. ТИПЫ СВЯЗЕЙ И ОБЩИЕ ЗАМЕЧАНИЯ О СВЯЗЯХ

В биологических явлениях часто наблюдаются связи между отдельными качественными или количественными признаками данного организма или между признаками и какими-либо внешними условиями. Так, например, всем хорошо известна связь между возрастом и весом тела, между высотой и весом организма, между пигментацией волос и цветом глаз, между температурой окружающего воздуха и частотой дыхания и т. п.

Связи биологических явлений или процессов, как правило, отличаются от явлений, наблюдаемых в неживой природе, с которыми имеют дело точные науки — физика, химия, математика.

Например, если взять несколько проб газа и изменить для них температуру с 20° до 25°C, то во всех пробах, находящихся в одинаковых условиях давления и измененной температуры, газ увеличит свой объём на определенную величину, одинаковую для всех взятых проб.

Если же осуществить аналогичные изменения температуры воздуха, окружающего группу животных, однородную по виду, породе, полу и возрасту, то реакция каждого животного на измененные температурные условия будет разная (если в качестве показателя взять, например, частоту дыхания). На повышение температуры животные в массе реагируют учащением дыхательных движений; при этом часть животных более значительно увеличит частоту дыхания, часть — менее, и будет какое-то число животных, очень мало изменивших частоту дыхательных движений. Таким образом, на фоне имеющейся связи между температурой окружающего воздуха и частотой дыхания у животных будет проявляться вариабильность этого показателя, зависящая от многих случайных причин.

Связи, приведенные в первом примере, имеют точно функциональный характер, т. е. при изменении одного показателя (температуры) другой (объём газа) изменяется во всех массовых пробах на одинаковую величину.

Во втором примере говорится о статистических связях, которые сопровождаются варьированием величин второго признака (частота дыхания) при определенном изменении первого признака (температуры).

Такие связи получили название корреляционных связей.

Наличие коррелятивной изменчивости у организмов было замечено биологами давно. Ч. Дарвин посвятил этому вопросу большую часть своих исследований, в основе которых лежат биологические приемы.

Но в то же время корреляционные связи могут быть изучены и с помощью вариационно-статистического метода.

Корреляционные связи могут быть прямыми и обратными, прямолинейными и криволинейными, простыми и множественными.

При прямом типе связи увеличение (или уменьшение) одного из коррелирующих признаков вызывает увеличение (или уменьшение) другого. Например, при увеличении возраста дерева увеличивается число колец (слоев) у ствола, или: с уменьшением окружающей температуры воздуха частота дыханий уменьшается.

При обратном типе связи с увеличением одного из коррелирующих признаков другой уменьшается. Так, при увеличении числа детенышей в помете крольчихи вес каждого крольчонка будет уменьшаться, или: с увеличением дозы облучения плодовитость самок уменьшается.

Обычно прямые и обратные связи имеют прямолинейный тип или близкий к нему (рис. 13 и 14).

Криволинейным типом связи называется такая связь, при которой с увеличением одного признака другой признак сначала увеличивается, а затем уменьшается, или, наоборот, с уменьшением одного признака другой сначала уменьшается, а затем увеличивается.

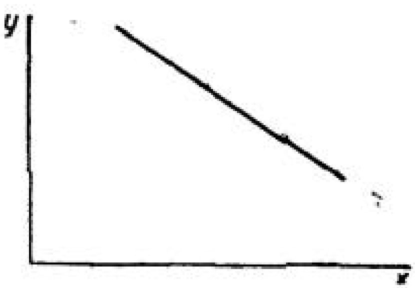

Рис. 13

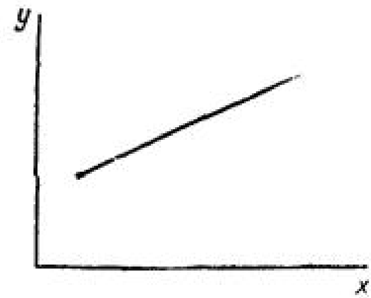

Рис. 14

Примером могут служить многие изменения в организме, имеющие возрастной характер. Так, с увеличением возраста коров лактационные удои увеличиваются до 5—8-го отёла, а затем постепенно уменьшаются (рис. 15*а*). На графике этот тип связи изображен в виде выпуклой кривой. Такого же криволинейного типа будет связь урожайности с количеством выпавших осадков и т. п.

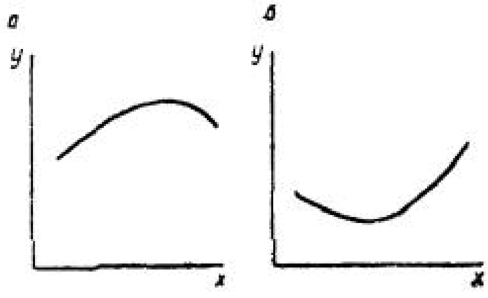

Рис. 15

Величина и характер связи могут быть выявлены с помощью вычисления статистических характеристик, среди которых наиболее распространены и используются такие, как коэффициент корреляции, коэффициент регрессии, корреляционное отношение и некоторые другие.

Кроме этого, можно найти уравнения, отражающие закономерность связи и позволяющие построить кривую связи.

Простыми случаями корреляционных связей будут такие, когда связь имеет место между двумя признаками, что было показано в приведенных выше примерах.

Но простая корреляция между двумя признаками в биологии, как правило, не встречается, так как обычно какой-либо признак у организма связан со многими факторами и другими признаками. Например, урожайность связана с уровнем осадков, удобрений, температуры, качеством сорта, агротехникой и т. п.

Поэтому, вычисление связи между двумя признаками осуществляется без учета других возможных связей, как бы абстрагируясь от их влияния. Это допущение можно делать, выделяя в качестве двух признаков ведущие и стараясь при организации выборки добиваться того, чтобы другие возможные факторы влияния на изучаемые признаки были одинаковыми для всех членов выборки.

Так, например, если изучается связь урожайности с температурными условиями, то стараются обеспечить для всех зёрен одинаковые условия по освещенности, влажности, сортности, агротехнике и т. п.

Вариационная статистика позволяет вычислять связи и между большим числом признаков (множественная корреляция), но при этом очень сильно усложняется вычислительная сторона, в связи с чем на практике чаще всего имеют дело с обработкой двух коррелирующих показателей.

Измеряя величину связи, можно вычислять ее как для одноименных, так и разноименных признаков (*см* и *кг*, *л* и возраст в годах и т. п.), и даже определять связь между количественными и качественными признаками (например, связь между возрастом родителей и полом новорожденных, между пигментацией кожного покрова и типом заболевания и т. п.).

При вычислении тех или иных коэффициентов связи следует помнить, что с помощью их только констатируется и измеряется величина связи, но не вскрывается причинность связей. Например, если определено наличие большой связи между показателями A и B, то эта связь может иметь разную основу: или A — причина изменения B, или B — причина изменения A, или есть еще фактор C, вызывающий изменения и A и B, сопутствующие друг другу, но зависящие от C.

2. КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ r И ТЕХНИКА ЕГО ВЫЧИСЛЕНИЯ ПРИ РАЗЛИЧНЫХ УСЛОВИЯХ

Коэффициент корреляции r является наиболее распространенной статистической характеристикой связи, используемой в биологии, и применяется в тех случаях, когда между признаками имеет место прямолинейная или близкая к этому типу связь. При криволинейной связи коэффициент корреляции преуменьшает величину связи и даже может ее не уловить.

Коэффициент корреляции представляет собой десятичную дробь, значение которой заключено в пределах от -1 до $+1$. Знак «минус» указывает на обратный тип связи, знак «плюс» говорит о положительном типе.

В общем виде формула коэффициента корреляции выглядит так:

$$r = \frac{\Sigma xy}{n},$$

где x — нормированное отклонение первого признака, т. е. $\frac{\Sigma(x-\overline{x})}{\sigma_x}$;

y — нормированное отклонение второго признака, т. е. $\frac{\Sigma(y-\overline{y})}{\sigma_y}$;

n — число особей, у которых изучаются признаки x и y.

Исходя из этого, формула может быть написана и в таком виде:

$$r = \frac{\Sigma\left(\frac{x-\overline{x}}{\sigma_x}\right)\cdot\left(\frac{y-\overline{y}}{\sigma_y}\right)}{n}.$$

Коэффициент корреляции можно представить в виде более простой формулы. Введем показатель ковариации (связи):

$$Co_{xy} = \frac{\Sigma p_{xy}(x-\overline{x})(y-\overline{y})}{n},$$

где p_{xy} обозначает частоту какой-либо пары признаков x и y; $(x-\overline{x})$ и $(y-\overline{y})$ — отклонения каждого признака от своей средней арифметической.

Запишем коэффициент корреляции, используя ковариацию:

$$r = \frac{Co_{xy}}{\sigma_x \cdot \sigma_y}.$$

Ковариацию можно упростить:

$$Co_{xy} = \frac{\Sigma p_{xy} \cdot xy}{n} - \overline{x} \cdot \overline{y}.$$

Средние квадратические отклонения σ_x и σ_y можно представить в виде, удобном для вычислений:

$$\sigma_x = \sqrt{\frac{\Sigma p_x \cdot x}{n} - \overline{x}^2}; \ \sigma_y = \sqrt{\frac{\Sigma p_y \cdot y}{n} - \overline{y}^2}.$$

В наших последующих примерах воспользуемся и другими выражениями формулы коэффициента корреляции, которые, не отличаясь от предыдущих по существу, дают также простые способы расчетов для определения r. Эти же приёмы были до некоторой степени использованы при изложении техники вычисления M и σ для малых и больших выборок (стр. 44).

а. Вычисление коэффициента корреляции для малой выборки при однозначных и двузначных значениях вариантов

Используем сумму произведения вариантов в формуле r:

$$r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\alpha_x \cdot \alpha_y}},$$

где x — варианты первого признака;
y — варианты второго признака;
n — число наблюдений;

$$\alpha_x = \Sigma x^2 - \frac{(\Sigma x)^2}{n}; \qquad \alpha_y = \Sigma y^2 - \frac{(\Sigma y)^2}{n}.$$

Пример. Определить величину и направление связи между числом рожденных мышат и дозой облучения их матерей во время беременности лучами рентгена (см. табл. на стр. 76).

$$\alpha_x = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 3527 - \frac{177^2}{10} = 3527 - \frac{31329}{10} =$$

$$= 3527 - 3132,9 = 394,1.$$

№ самки	Доза облучения в рентгенах, x	Число рожденных мышат, y	x^2	y^2	xy
1	10	10	100	100	100
2	15	9	225	81	135
3	20	5	400	25	100
4	10	9	100	81	90
5	12	7	144	49	84
6	15	6	225	36	90
7	18	8	324	64	144
8	22	3	484	9	66
9	25	4	625	16	100
10	30	1	900	1	30
	$\Sigma x = 177$	$\Sigma y = 62$	$\Sigma x^2 = 3527$	$\Sigma y^2 = 462$	$\Sigma xy = 939$

Отсюда:

$$\sigma_x = \sqrt{\frac{\alpha}{n-1}} = \sqrt{\frac{394,1}{10-1}} = \sqrt{43,8} = 6,618 \text{ рентген.}$$

$$\alpha_y = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 462 - \frac{62^2}{10} = 462 - \frac{3844}{10} = 77,6.$$

$$\sigma_y = \sqrt{\frac{\alpha}{n-1}} = \sqrt{\frac{77,6}{10-1}} = \sqrt{8,62} = 2,936 \text{ голов.}$$

Отсюда:

$$r = \frac{\Sigma xy - \frac{\Sigma x \cdot \Sigma y}{n}}{\sqrt{\alpha_x \cdot \alpha_y}} = \frac{939 - \frac{177 \cdot 62}{10}}{\sqrt{394,1 \cdot 77,6}} = \frac{939 - 1097,4}{\sqrt{30582,16}} =$$

$$= -\frac{158,4}{174,8} = -0,906.$$

Делаем заключение, что связь плодовитости с дозой облучения в пределах от 10 до 30 рентген велика и имеет обратное направление, т. е. чем выше доза облучения, тем меньше плодовитость самок.

б. Вычисление коэффициента корреляции при многозначных или дробных значениях вариантов

Для упрощения расчетов выразим варианты через их отклонение Δ от условной средней A. Рабочая формула r будет такова:

$$r = \frac{\Sigma\Delta_x \cdot \Delta_y - \frac{\Sigma\Delta_x \cdot \Sigma\Delta_y}{n}}{\sqrt{\alpha_x \cdot \alpha_y}},$$

где

$$\alpha_x = \Sigma\Delta_x^2 - \frac{(\Sigma\Delta_x)^2}{n}; \quad \alpha_y = \Sigma\Delta_y^2 - \frac{(\Sigma\Delta_y)^2}{n}.$$

Так же, как и для вычисления σ, о чем было сказано на стр. 44, требуется определить условные средние A_x и A_y, взять разницу Δ между каждым вариантом и соответствующей величиной A_x и A_y, т. е. $(x - A_x)$ и $(y - A_y)$ и затем определить α_x и α_y.

Пример. Определить связь между живым весом и удоем за лактационный период у коров остфризской породы по первой лактации (из книги Н. А. Плохинского).

Вес в *кг*, *x*	Удой за лактационный период в *кг*, *y*	$\Delta_x = x - A_x$	$\Delta_y = y - A_y$	Δ_x^2	Δ_y^2	$\Delta_x \cdot \Delta_y$
450	2020	450—450=0	2020—2040=—20	0	400	0
420	2030	—30	—10	900	100	+ 300
470	2050	+20	+10	400	100	+ 200
400	2000	—50	—40	2500	1600	+2000
490	2090	+40	+50	1600	2500	+2000
410	2010	—40	—30	1600	900	+1200
480	2060	+30	+20	900	400	+ 600
440	2070	—10	+30	100	900	— 300
460	2040	+10	0	100	0	0
430	2080	—20	+40	400	1600	— 800
$A_1 = 450$ $A_2 = 2040$		$\Sigma\Delta_x = -50$	$\Sigma\Delta_y = +50$	$\Sigma\Delta_x^2 = 8500$	$\Sigma\Delta_y^2 = 8500$	$\Sigma\Delta_x \cdot \Delta_y = 5200$

$$\alpha_x = \Sigma\Delta_x^2 - \frac{(\Sigma\Delta_x)^2}{n} = 8500 - 250 = 8250;$$

$$\sigma_y = \Sigma\Delta_y^2 - \frac{(\Sigma\Delta_y)^2}{n} = 8500 - 250 = 8250;$$

$$r = \frac{\Sigma\Delta_x \cdot \Delta_y - \frac{\Sigma\Delta_x \cdot \Sigma\Delta_y}{n}}{\sqrt{\alpha_x \cdot \alpha_y}} = \frac{5200 - \frac{-2500}{10}}{\sqrt{8250 \cdot 8250}} = 0{,}66.$$

Из примера видно, что связь значительная (r=0,66) и с увеличением веса коров увеличивается и их удой.

в. Вычисление коэффициента корреляции для альтернативных признаков

Если необходимо определить связь между альтернативными признаками, то пользуются следующей формулой коэффициента корреляции:

$$r = \frac{p_1 \cdot p_4 - p_2 \cdot p_3}{\sqrt{(p_1 + p_2)(p_3 + p_4)(p_1 + p_3)(p_2 + p_4)}}.$$

Здесь p_1, p_2, p_3, p_4 — число наблюдений в каждой клетке корреляционной решётки по обоим признакам x и y.

Пример. Определить, имеется ли связь между типом осеменения родительского поколения подопытных кур и наследованием пигментации ног их потомством.

Размещаем наши данные в так называемую корреляционную решётку, где классами будут служить интересующие нас признаки (табл. 7).

Таблица 7

Корреляционная решётка для качественных признаков

Окраска ног, y	Тип осеменения, x	
	парное скрещивание	осеменение смесью семени
Белая	49 (p_1)	20 (p_2)
Черная	1 (p_3)	30 (p_4)
Σp	$p_1 + p_3 = 50$	$p_2 + p_4 = 50$

$$r = \frac{p_1 \cdot p_4 - p_2 \cdot p_3}{\sqrt{(p_1+p_2)(p_3+p_4)(p_1+p_3)(p_2+p_4)}} = \frac{49 \cdot 30 - 20 \cdot 1}{\sqrt{69 \cdot 31 \cdot 50 \cdot 50}} =$$

$$= \frac{1450}{\sqrt{5347500}} = \frac{1450}{2312,1} = 0,627;$$

т. е. связь между типом осеменения и пигментацией ног у кур значительная ($r=0,627$).

г. Определение коэффициента корреляции между количественными признаками при большом числе наблюдений

Для определения связи между двумя признаками на большом числе наблюдений составляется корреляционная решётка, в которой по горизонтали указаны классы по одному признаку, а по вертикали — классы по второму признаку. Принцип разбивки ряда на классы таков же, какой использовался при составлении вариационных рядов для вычисления M и σ (см. стр. 33).

После составления классов по каждому из коррелирующих признаков делается разноска членов выборки по клеткам решётки. Дальнейшая обработка корреляционной решётки осуществляется так, чтобы были получены необходимые величины, входящие в формулу.

Формула коэффициента корреляции для большой выборки выглядит так:

$$r = \frac{\Sigma p a_x \cdot a_y - n b_x \cdot b_y}{n \sigma_x \cdot \sigma_y}.$$

Отсюда видно, как следует вести обработку корреляционной решётки.

Напомним, что $b_x = m_{1(x)} = \frac{\Sigma p_x \cdot a_x}{n}$; $b_y = m_{1(y)} = \frac{\Sigma p_y a_y}{n}$, n — число наблюдений;

$$\sigma_x = \sqrt{\frac{\Sigma\, p_x \cdot a_x^2}{n} - m_{1(x)}^2}; \quad \sigma_y = \sqrt{\frac{\Sigma\, p_y \cdot a_y^2}{n} - m_{1(y)}^2}$$

Отметим, что значение σ для коэффициента корреляции дается в относительных величинах, а не в именованных показателях признака, т. е. без умножения величины σ на классовый промежуток k. Поэтому-то и можно в знаменателе делать умножение значений изменчивости обоих признаков, выраженной в относительных величинах.

Величины $b_x = m_{1_x}$ и $b_y = m_{1_y}$ определяются известным нам способом, приведенным на стр. 35 для каждого признака. Произведение $pa_x \cdot a_y$ получается путем умножения частот, размещенных по клеткам корреляционной решётки, на соответствующие им условные отклонения по обоим признакам a_x и a_y.

Для вычисления всех членов формулы (b_x, b_y, σ_x, σ_y) нужно по каждому признаку выделить класс условной средней A_x и A_y. Указанная обработка корреляционной решётки позволяет одновременно вычислять не только коэффициент корреляции, но также M_x и M_y, σ_x и σ_y, C_x и C_y, т. е. она дает и необходимые характеристики каждого вариационного ряда, из которых образована корреляционная решётка.

Пример. Найти величину и направление связи r между удоем за лактационный период и жирномолочностью коров, выраженной показателем среднего процента жира за лактацию. Одновременно вычислить M, σ и C для удоя и жирномолочности.

1. Найдем размах колебания каждого признака и определим величину классового промежутка:

удой в *кг*, x:

$$\begin{array}{r} -\begin{array}{l} 4500 \text{ (макс)} \\ 2300 \text{ (мин)} \end{array} \\ \hline D_x = 2200 \end{array} \qquad k_x = \frac{D_x}{10} = \frac{2200}{10} = 220 \text{ кг};$$

жирномолочность в %, y:

$$\begin{array}{r} -\begin{array}{l} 4{,}5 \text{ (макс)} \\ 2{,}7 \text{ (мин)} \end{array} \\ \hline D_y = 1{,}8 \end{array} \qquad k_y = \frac{1{,}8}{10} = 0{,}18\%.$$

Округляем значение классов k:

$$k_x = 200 \text{ кг}, \quad k_y = 0{,}2\%.$$

2. Составляем классы и строим корреляционную решётку (см. табл. 8).

3. Делаем разноску материала по клеткам решётки методом квадратов $\boxed{\times}$.

4. Подсчитываем частоты по каждому классу признаков (p_x и p_y) и получаем вариационные ряды по обоим признакам.

5. Выделяем классы условных средних A_x и A_y, по воз-

Таблица 8

Корреляционная решётка для удоев и жирномолочности

Удой в *кг*, x	% жира, y													
	2,7—2,89	2,9—3,09	3,1—3,29	3,3—3,49	3,5—3,69 A_y	3,7—3,89	3,9—4,09	4,1—4,29	4,3—4,49	4,5—4,69	p_x	a_x	$p_x a_x$	$p_x a_x^2$
	I квадрант					II квадрант								
2300—2499	—	—	—	—	—	—	—	—	1	1	2	−5	−10	50
2500—2699	—	—	—	—	—	—	1	1	1	—	3	−4	−12	48
2700—2899	—	—	—	—	—	—	3	—	2	—	5	−3	−15	45
2900—3099	—	—	—	—	1	—	3	—	—	—	4	−2	−8	16
3100—3299	—	—	—	—	5	3	2	—	—	—	10	−1	−10	10
A_x 3300—3499	—	—	—	—	32	5	2	—	—	—	39	0	—	—
	III квадрант					IV квадрант								
3500—3699	—	—	—	3	7	3	2	—	—	—	15	+1	15	15
3700—3899	—	—	1	2	5	2	—	—	—	—	10	+2	20	40
3900—4099	—	—	2	2	—	1	1	—	—	—	6	+3	18	54
4100—4299	—	—	2	1	—	—	—	—	—	—	3	+4	12	48
4300—4499	—	2	—	—	—	—	—	—	—	—	2	+5	10	50
4500—4699	1	—	—	—	—	—	—	—	—	—	1	+6	6	36
p_y	1	2	5	8	50	14	14	1	4	1	$n = 100$		$\Sigma p_x \cdot a_x = 26$	$\Sigma p_x \cdot a_x^2 = 412$
a_y	−4	−3	−2	−1	0	+1	+2	+3	+4	+5				
$p_y a_y$	−4	−6	−10	−8	0	14	28	3	16	5	$\Sigma p_y \cdot a_y = 38$			
$p_y a_y^2$	16	18	20	8	0	14	56	9	64	25	$\Sigma p_y \cdot a_y^2 = 230$			

можности занимающие центральные места в вариационных рядах и имеющие наибольшее число наблюдений n.

6. Выделяем классы условных средних жирными чертами и принимаем эти классы за нулевые, в результате чего в корреляционной решётке образуется «крест». Углы этого «креста» называются квадрантами.

7. Записываем условные отклонения a_x и a_y пс каждому признаку.

8. Вычисляем произведения pa и pa^2 и их суммы для каждого признака.

9. На основании проделанного из нашего примера получаем:

$$\Sigma p_x \cdot a_x = -55 + 81 = 26; \; b_x = \frac{\Sigma p_x \cdot a_x}{n} = \frac{26}{100} = 0{,}26;$$

$$\Sigma p_y \cdot a_y = -28 + 66 = 38; \; b_x^2 = 0{,}26^2 = 0{,}0676;$$

$$\Sigma p_x \cdot a_x^2 = 412; \quad b_y = \frac{\Sigma p_y \cdot a_y}{n} = \frac{38}{100} = 0{,}38;$$

$$\Sigma p_y a_y^2 = 230; \quad b_y^2 = 0{,}38^2 = 0{,}1444;$$

$$A_x = 3300 + \frac{200}{2} = 3400 \text{ кг};$$

$$A_y = 3{,}5 + \frac{0{,}2}{2} = 3{,}6\% .$$

10. Определяем по каждому квадранту решётки (I, II, III и IV) выражение $\Sigma pa_x a_y$. I и IV квадранты дадут выражение $\Sigma pa_x \cdot a_y$ со знаком плюс, а II и III — со знаком минус. Получаем алгебраическую сумму по всем квадрантам, которая и войдет в формулу. Следует иметь в виду, что в нулевом классе (класс условной средней A) $pa_x a_y$ дает нуль.

Для нашего примера:

I квадрант $pa_x \cdot a_y$	IV квадрант $pa_x \cdot a_y$	II квадрант $pa_x \cdot a_y$	III квадрант $p \cdot a_x \cdot a_y$
	$3 \cdot 1 \cdot 1 = 3$	$1\,(-5)\,4 = -20$	$3 \cdot 1\,(-1) = -3$
	$2 \cdot 1 \cdot 2 = 4$	$1\,(-5)\,5 = -25$	$1 \cdot 2\,(-2) = -4$
	$2 \cdot 2 \cdot 1 = 4$	$1\,(-4)\,2 = -8$	$2 \cdot 2\,(-1) = -4$
	$1 \cdot 3 \cdot 1 = 3$	$1\,(-4)\,3 = -12$	$2 \cdot 3\,(-2) = -12$
	$1 \cdot 3 \cdot 2 = 6$	$1\,(-4)\,4 = -16$	$2 \cdot 3\,(-1) = -6$
		$3\,(-3)\,2 = -18$	$2 \cdot 4\,(-2) = -16$
		$2\,(-3)\,4 = -24$	$1 \cdot 4\,(-1) = -4$
		$3\,(-2)\,2 = -12$	$2 \cdot 5\,(-3) = -30$
		$3\,(-1)\,1 = -3$	$1 \cdot 6\,(-4) = -24$
		$2\,(-1)\,2 = -4$	
$\Sigma\, pa_x \cdot a_y = 0$	$\Sigma\, pa_x \cdot a_y = 20$	$\Sigma\, pa_x \cdot a_y = -142$	$\Sigma\, pa_x \cdot a_y = -103$

Отсюда общая алгебраическая сумма по четырем квадрантам равна: $\Sigma pa_x \cdot a_y = 0 + 20 - 142 - 103 = -225$.

11. Вычислим средние квадратические отклонения σ_x и σ_y:

$$\sigma_x = k_x \sqrt{\frac{\Sigma p_x a_x^2}{n} - b_x^2} = 200 \sqrt{\frac{412}{100} - 0{,}0676} =$$

$$= 200 \sqrt{4{,}0524} = 200 \cdot 2{,}01 = 402 \text{ кг};$$

$$\sigma_y = k_y \sqrt{\frac{\Sigma p_y \cdot a_y^2}{n} - b_y^2} = 0{,}2 \sqrt{\frac{230}{100} - 0{,}1444} =$$

$$= 0{,}2 \sqrt{2{,}3 - 0{,}1444} = 0{,}2 \sqrt{2{,}1556} = 0{,}2 \cdot 1{,}468 = 0{,}29\%.$$

12. Подставляем полученные значения в формулу коэффициента корреляции:

$$r = \frac{\Sigma pa_x \cdot a_y - nb_x \cdot b_y}{n\sigma_x \cdot \sigma_y} = \frac{-225 - 100 \cdot 0{,}26 \cdot 0{,}38}{100 \cdot 2{,}01 \cdot 1{,}468} =$$

$$= \frac{-225 - 9{,}88}{295{,}068} = -\frac{215{,}12}{295{,}068} = -0{,}7324.$$

Таким образом, связь между удоем за лактационный период и жирномолочностью коров оказалась высокой и обратной, т. е. увеличение удоя снижает жирномолочность.

Из этого статистического показателя можно сделать вывод: если вести массовую селекцию в стаде на повышение удоя, то это будет сопровождаться некоторым снижением содержания жира в молоке коров. Таким образом, статистический анализ подводит нас к заключению, имеющему практическое селекционное значение.

Из приведенного примера видно, что статистический анализ, констатируя величину и уровень связи, служит подсобным анализом для селекционной работы и указывает на необходимость разработки мероприятий, направленных на преодоление нежелательных для человека коррелятивных связей в организме животного.

13. Ранее отмечалось, что обработка корреляционной решётки позволяет вычислить M, σ и C. Осуществим это вычисление:

$$M = A + bk;$$

$$M_x = A_x + b_x k_x = 3400 + 0{,}26 \cdot 200 = 3452 \text{ кг};$$

$$M_y = A_y + b_y k_y = 3{,}6 + 0{,}38 \cdot 0{,}2 = 3{,}67\%;$$

$$C_x = \frac{\sigma_x}{M_x} \cdot 100 = \frac{402}{3452} \cdot 100 = 11{,}6\%;$$

$$C_y = \frac{\sigma_y}{M_y} \cdot 100 = \frac{0{,}29}{3{,}67} \cdot 100 = 7{,}9\%.$$

По коэффициенту изменчивости C видно, что удой — более вариабильный признак, чем жирномолочность. Это подтверждает и практика. Известно, что под влиянием, например, кормления удой коров значительно меняется в сторону его увеличения или уменьшения, а процент жира в молоке при этом остается почти без изменения.

Из показателя коэффициента изменчивости можно сделать селекционный вывод: закрепить селекцией легче жирномолочность, как менее вариабильный признак, и труднее — удой.

В то же время известно, что при большой вариабильности признака материал для отбора является более широким, чем у признака с меньшей вариабильностью, что имеет практическое значение для расширения отбора.

Такая генетическая противоречивость между изменчивостью и константностью свойств часто сопровождает работу селекционера. Статистический же анализ, дающий показатели σ, M и C, позволяет правильно вскрыть эти особенности и наметить правильное решение в направлении селекционной работы.

По расположению частот в корреляционной решётке можно заранее, до ее обработки, предвидеть направление и приблизительную величину связи. Рассмотрим рис. 16 *а, б, в, г:*

1) если частоты имеют тенденцию размещаться по всей решётке, то связь между признаками отсутствует (рис. 16*а*);

2) если частоты располагаются ближе к диагонали из нижнего левого к правому верхнему углу, то связь значительная и обратная (рис. 16*б*);

3) если частоты располагаются по диагонали, идущей от правого нижнего к левому верхнему углу, то связь значительная и прямая (рис. 16*в*);

4) если частоты располагаются в виде подковы, то связь имеет криволинейный тип, и вычисление ее величины с помо-

щью коэффициента корреляции невозможно; в этом случае следует пользоваться другим коэффициентом связи (рис. 16г).

Связь между признаками можно выразить не только с помощью коэффициента корреляции, но и через уравнение прямой. Если рассматривается прямолинейная связь между дву-

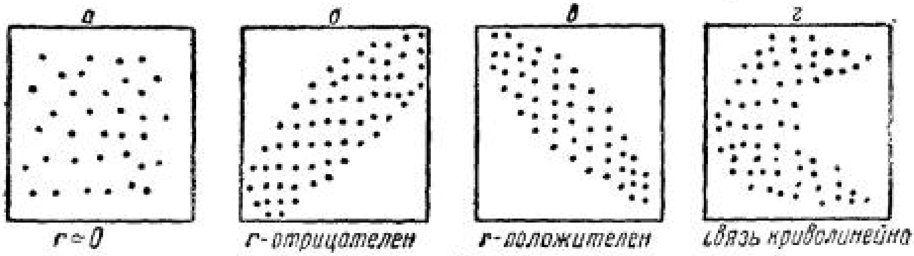

Рис. 16

мя признаками, то в общей форме ее можно представить уравнением прямой $y=a+bx$, где x и y — показатели коррелирующих вариантов; a и b — параметры, по которым можно выявить особенности прямой, отражающей связь между x и y. Для решения уравнений прямой используется способ наименьших квадратов.

3. РЕГРЕССИЯ КАК ПОКАЗАТЕЛЬ СВЯЗИ R

(уравнение регрессии, коэффициент регрессии, линия регрессии)

Регрессия показывает величину изменения одного признака при изменении другого на определенную величину.

Так, например, можно вычислить, насколько уменьшается плодовитость мышей при облучении их той или иной дозой лучей рентгена. По линии регрессии можно для любого значения дозы (x) найти соответствующую ей плодовитость самок (y), несмотря на то, что в данном опыте эта доза и не употреблялась.

Можно ставить вопрос и в обратном смысле, а именно — какому значению y будет соответствовать определенное значение x. Так, например, можно определить, какая доза облучения (x) будет соответствовать определенной плодовитости (y).

Уравнение прямой $y=a+bx$ называется также уравнением регрессии, а построенный на основе этого уравнения график для конкретных данных называется эмпирической линией регрессии.

Для выяснения зависимости y при определенном значении x или зависимости x от определенного значения y пользуются коэффициентом регрессии R, формула которого выглядит так:

$$R_{x\,y} = r\frac{\sigma_x}{\sigma_y} \text{ или } R_{y\,x} = r\cdot\frac{\sigma_y}{\sigma_x}.$$

Коэффициент регрессии в графическом виде отвечает тангенсу угла наклона прямой, изображающей зависимость, к оси абсцисс.

В формулах регрессии r означает коэффициент корреляции; σ_x и σ_y — средние квадратические отклонения по признакам x и y. Из приведенных формул видно, что коэффициент регрессии для одного и того же ряда имеет два значения: $R_{x\,y}$ означает, что определяется величина x при определенном значении y; $R_{y/x}$ показывает обратное, т. е. какова будет величина y при определенном значении x.

Разберем эти особенности коэффициента регрессии на примере с плодовитостью мышей при разных дозах облучения. На стр. 77 было найдено, что связь между этими показателями выражается коэффициентом корреляции, равным — 0,906. При этом средние квадратические отклонения были следующие: для показателя плодовитости $\sigma_y = 2{,}94$ голов и для дозы облучения $\sigma_x = 6{,}62$ рентгена.

Подставляя в формулу коэффициента регрессии эти выражения, получим:

$$R_{y\,x} = -0{,}906\cdot\frac{2{,}94}{6{,}62} = -\frac{2{,}664}{6{,}62} = -0{,}402 \approx -0{,}4 \text{ головы.}$$

При изменении облучения на единицу плодовитость уменьшается на 0,4 головы.

Вычислим второе значение регрессии:

$$R_{x\,y} = -0{,}906\cdot\frac{6{,}62}{2{,}94} = -\frac{5{,}997}{2{,}94} = -2{,}039 \approx -2{,}04 \text{ рентгена.}$$

Это значение регрессии показывает, что для уменьшения плодовитости в среднем на 1 голову в помете облучение должно быть увеличено на 2,04 рентгена.

Использование коэффициента регрессии особенно удобно при необходимости осуществить некоторый прогноз в отношении одного показателя на основе информации о другом.

Пример. Связь между удоем и питательностью рациона выражается коэффициентом корреляции $r=0{,}8$. Изменчивость суточного удоя выражается $\sigma_x=2$ *кг*, а изменчивость питательности рациона — $\sigma_y=0{,}9$ кормовых единиц. Определить:

а) На сколько килограмм увеличится суточный удой коро-

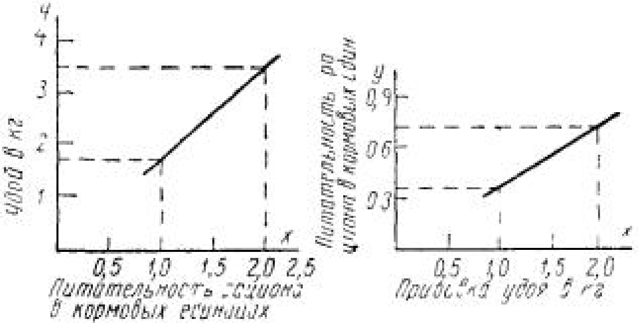

Рис. 17 и 18

вы при увеличении питательности рациона на 1 кормовую единицу?

Подставляем имеющиеся данные в формулу коэффициента регрессии, получаем:

$$R_{xy}=r\cdot\frac{\sigma_x}{\sigma_y}=0{,}8\cdot\frac{2}{0{,}9}=\frac{1{,}6}{0{,}9}=1{,}77\ \text{кг},$$

т. е. удой увеличится на 1,77 *кг* молока, если питательность рациона увеличится на 1 кормовую единицу.

б) Сколько потребуется добавить кормовых единиц в рационе, чтобы увеличить удой на 1 *кг* молока?

Расчет дает следующее:

$$R_{y|x}=r\,\frac{\sigma_y}{\sigma_x}=0{,}8\cdot\frac{0{,}9}{2}=\frac{0{,}72}{2}=0{,}36\ \text{кормовых единиц},$$

т. е. на каждый килограмм молока следует добавлять 0,36 кормовых единиц питательных веществ.

По коэффициенту регрессии можно построить графическое изображение эмпирической линии регрессии.

На рис. 17 и 18 приведены графики линий регрессии, построенные на основе данных примера.

Так как коэффициенты корреляции и регрессии отражают связь при прямолинейном ее типе, то можно по уравнению прямой и по показателю регрессии определить значения зависимой переменной при изменении на одинаковые отрезки значений независимой переменной.

Уравнение прямой имеет вид $y=a+bx$, но оно может быть представлено также через коэффициент регрессии в следующем общем виде: $y=a+Rx$, что можно выразить в развернутом виде и так: $y=a+r\frac{\sigma_x}{\sigma_y}x$. Но это уравнение подразумевает, что начало координат проходит через точку, соответствующую значению средней арифметической. Так как это не всегда имеет место, то лучше значения y и x выразить через их отклонения от своих средних арифметических $\overline{y}$ и $\overline{x}$, тогда уравнение может быть составлено в следующей форме:

$$y-\overline{y}=r\cdot\frac{\sigma_y}{\sigma_x}\cdot(x-\overline{x}),$$

откуда находим y:

$$y=\overline{y}+r\cdot\frac{\sigma_y}{\sigma_x}\cdot(x-\overline{x}).$$

Если требуется определить неизвестное x по значениям y, то уравнение соответственно видоизменяется так:

$$x-\overline{x}=r\cdot\frac{\sigma_x}{\sigma_y}\,.\,(y-\overline{y}).$$

Отсюда:

$$x=\overline{x}+r\,.\,\frac{\sigma_x}{\sigma_y}\,.\,(y-\overline{y}).$$

На основании этих уравнений, зная средние арифметические обоих признаков, среднеквадратические отклонения и коэффициент корреляции, можно получить коэффициенты регрессии x по y и y по x и построить две эмпирические линии регрессии.

Пусть из примера на стр. 88 известно, что средний суточный удой $\overline{x}=20$ *кг*; $\sigma_x=2$ *кг*; $\sigma_y=0{,}9$ кормовых единиц; средняя питательность рациона $\overline{y}=7$ кормовых единиц; $r=0{,}80$; $R_{x/y}=1{,}77$ *кг*; $R_{y/x}=0{,}36$ кормовых единиц.

Определить, как будет изменяться удой при различных значениях питательности рациона.

Подставляем различные значения питательности рациона y (аргумент) в уравнение прямой и находим соответствующие значения удоя x (функция) (буквы в уравнении для обозначения функции и аргумента берем в соответствии со смыслом и обозначениями в нашем примере):

$$x - \bar{x} = R_{xy}(y - \bar{y});$$

$$x - 20 = 1{,}77\,(y - 7);$$

откуда:

$$x = 1{,}77y - 12{,}39 + 20$$

или:

$$x = 1{,}77y + 7{,}61.$$

Если в последнее уравнение подставим значения питательности рациона, то получим данные об изменении x (удоя за сутки) при разном уровне питания. Так, если питательность рациона y будет, например, равна 5, 10, 15 кормовым единицам, то удой x изменится так:

$y = 5$ кормовым единицам, тогда $x = 1{,}77 \cdot 5 + 7{,}61 = 16{,}46$ *кг*,
$y = 10$ кормовым единицам, тогда $x = 1{,}77 \cdot 10 + 7{,}61 = 25{,}31$ *кг*,
$y = 15$ кормовым единицам, тогда $x = 1{,}77 \cdot 15 + 7{,}61 = 33{,}96$ *кг*.

На этом принципе в зоотехнической литературе составлены расчеты норм кормления в зависимости от уровня продуктивности (продуктивный корм) или от величины живого веса (поддерживающий корм).

Таким образом, зная $\bar{x}$, $\bar{y}$, σ, r и R, можно планировать изменения показателя функции при изменении показателя аргумента на определенную величину (например, планирование привесов и живого веса от питательности рациона, урожайности от удобрения или температурных условий, плодовитости от облучения, кормления или других факторов).

До сих пор нами рассматривались самые простые формы связи, а именно, прямолинейная связь между двумя показателями. Но чаще имеет место более сложная связь, которая обнаруживается между несколькими коррелирующими признаками. Такой тип связи называется множественной корреляцией. Примером этой связи может быть зависимость урожая от осадков, температуры и удобрения или зависимость удоя от возраста, кормления, породы и т. п.

Изучение этих связей осуществляется методом множественной корреляции, который в данном пособии не рассматривается.

4. ПОЛИХОРИЧЕСКИЙ ПОКАЗАТЕЛЬ СВЯЗИ ρ
(второй коэффициент сопряженности Чупрова)

При определении связи между качественными признаками, величина которых определяется грубо на глаз и когда число классов больше двух, можно пользоваться так называемым полихорическим показателем связи ρ, имеющим всегда положительное значение в пределах от 0 до 1. Направление связи определяется по характеру распределения частот в корреляционной решётке.

Полихорический показатель связи выражается следующей формулой:

$$\rho = \frac{\alpha - 1}{\sqrt{(l_1 - 1)(l_2 - 1)}},$$

где $(\alpha-1)$ — так называемый «показатель контингенции» Пирсона;

l_1 и l_2 — число классов по каждому признаку.

В формуле полихорического показателя связи

$$\alpha = \Sigma\left\{\frac{\Sigma p_{1\,2}^2 \cdot p_2}{p_1}\right\} - \frac{(l_1 - 1)(l_2 - 1)}{n}$$

$p_{1\,2}$ — частоты корреляционной решётки;
p_1 и p_2 — частоты ряда по каждому признаку;
n — число наблюдений;
l_1 и l_2 — число классов по каждому признаку.

Пример. Определить, имеется ли связь между типом облучения (рентгеновские лучи и поток тепловых нейтронов), которому подвергались семена ячменя, и типом мутаций. Корреляционная решётка оформлена в таблице (см. табл. 9).

Таблица 9

Корреляционная решётка по типу облучений и типу мутаций

Тип облучения, p_1	Тип мутаций, p_2						
	альбиносы	желтые	желто-зеленые и виресценс	полосатые	сморщенные	прочие	итого p_1
Рентгеновские лучи . . .	387	102	124	17	20	41	691
Поток тепловых нейтронов	779	140	279	36	52	86	1372
Итого по p_2 . . .	1166	242	403	53	72	127	2063

Из этой таблицы видно, что $n=2063$; число классов по первому признаку (тип облучения) $l_1=2$ и по второму (тип мутации) $l_2=6$.

Вычислим необходимые данные для значения α, т. е. для каждой частоты найдем в таблице значение $p_{1\cdot 2}^2$, разделим его на соответствующее значение p_2 и просуммируем все полученные значения $\frac{p_{1\cdot 2}^2}{p_2}$. Эти расчеты дают следующее:

по первой строчке:

$$\Sigma \frac{p_{1\cdot 2}^2}{p_2} = \frac{387^2}{1166} + \frac{102^2}{242} + \frac{124^2}{403} + \frac{17^2}{53} + \frac{20^2}{72} + \frac{41^2}{127} = 233{,}69;$$

по второй строчке:

$$\Sigma \frac{p_{1\cdot 2}^2}{p_2} = \frac{779^2}{1166} + \frac{140^2}{242} + \frac{279^2}{403} + \frac{36^2}{53} + \frac{52^2}{72} + \frac{86^2}{127} = 913{,}92.$$

Находим отсюда выражение:

$$\Sigma \left\{ \frac{\Sigma \frac{p_{1\cdot 2}^2}{p_2}}{p_1} \right\} = \frac{233{,}69}{691} + \frac{913{,}92}{1372} = 0{,}3381 + 0{,}6661 = 1{,}0042.$$

Подставляем эту величину в формулу α:

$$\alpha = \Sigma \left\{ \frac{\Sigma \frac{p_{1\cdot 2}^2}{p_2}}{p_1} \right\} - \frac{(l_1-1)(l_2-1)}{n} =$$

$$= 1{,}0042 - \frac{5}{2063} = 1{,}0018.$$

Делаем окончательный расчет, вычисляя значение полихорического коэффициента:

$$\rho = \frac{\alpha - 1}{\sqrt{(l_1-1)(l_2-1)}} = \frac{1{,}0018 - 1}{\sqrt{(2-1)(6-1)}} = \frac{0{,}0018}{\sqrt{5}} =$$

$$= \frac{0{,}0018}{2{,}236} \approx 0{,}001.$$

Таким образом, значение ρ очень мало; это указывает на отсутствие связи между типом облучения и типом полученных мутаций у ячменя.

Полихорический показатель связи можно применять и в том случае, если один из признаков измеряется грубо на глаз или принимает качественные выражения, а другой измеряется точно.

Пример. Найти связь между типом опыления и *озерненностью* колосьев ржи (табл. 10).

Таблица 10

Корреляционная решётка между типом опыления и степенью озернённости

Тип опыления, p_1	Число зёрен в колосе, p_2								
	0—4	5—9	10—14	15—19	20—24	25—29	30—34	35—39	p_1
Инцухт	15	25	20	—	—	—	—	—	60
Инцухт + чужеродная пыльца	—	5	30	10	10	5	—	—	60
Свободное опыление . . .	—	—	5	10	20	10	10	5	60
p_2	15	30	55	20	30	15	10	5	180

Известно, что

$$\alpha = \Sigma\left\{\frac{\Sigma\dfrac{p_{1\,2}^2}{p_2}}{p_1}\right\} - \frac{(l_1-1)(l_2-1)}{n}.$$

Находим для значения α необходимые данные по каждой строке:

$$\Sigma\frac{p_{1\,2}^2}{p_2} = \frac{15^2}{15} + \frac{25^2}{30} + \frac{20^2}{55} = 43,1;$$

$$\Sigma\frac{p_{1\,2}^2}{p_2} = \frac{5^2}{30} + \frac{30^2}{55} + \frac{10^2}{20} + \frac{10^2}{30} + \frac{5^2}{15} = 27,195;$$

$$\Sigma\frac{p_{1\,2}^2}{p_2} = \frac{5^2}{55} + \frac{10^2}{20} + \frac{20^2}{30} + \frac{10^2}{15} + \frac{10^2}{10} + \frac{5^2}{5} = 40,45;$$

$$\Sigma\left\{\frac{\Sigma\dfrac{p_{1\,2}^2}{p_2}}{p_1}\right\} = \frac{43,1}{60} + \frac{27,195}{60} + \frac{40,45}{60} =$$

$$= 0,72 + 0,42 + 0,67 = 1,84;$$

откуда:

$$\alpha = 1{,}84 - \frac{(3-1)(8-1)}{180} = 1{,}84 - 0{,}078 = 1{,}762.$$

Подставляем это значение α в основную формулу ρ:

$$\rho = \frac{1{,}762 - 1}{\sqrt{(3-1)(8-1)}} = \frac{0{,}762}{\sqrt{14}} \approx 0{,}201,$$

т. е. связь озернённости с типом опыления имеется, но она не велика.

5. БИСЕРИАЛЬНЫЙ ПОКАЗАТЕЛЬ СВЯЗИ r

В тех случаях когда один признак измеряется точно и может принять любое качественное значение, а другой может принять только два крайних (альтернативных) выражения, для вычисления связи между ними пользуются бисериальным показателем связи:

$$r_b = \frac{\sum \frac{p_+ \cdot a}{n_+} - \frac{\Sigma pa}{n}}{\sqrt{\frac{\alpha}{n_+} - \frac{\alpha}{n}}} \quad \text{или} \quad r_b = \frac{\sum \frac{p_- \cdot a}{n_-} - \frac{\Sigma pa}{n}}{\sqrt{\frac{\alpha}{n_-} - \frac{\alpha}{n}}},$$

где

$$\alpha = \Sigma pa^2 - \frac{(\Sigma pa)^2}{n};$$

p_+ — частоты первого альтернативного признака;
p_- — частоты второго альтернативного признака;
n — число наблюдений по всему ряду;
n_+ — число наблюдений по первому альтернативному признаку.

Для определения бисериального показателя связи обычным путем составляется корреляционная решётка. По вариантам количественного признака выделяется класс с условной средней A. Затем, как это делалось ранее, получают ряд условных отклонений a от класса с условной средней A. После этого вычисляются ряды pa и pa^2, что так же уже известно из материала, изложенного на стр. 82. Новой системой расчета для бисериального показателя является получение ряда, составленного из умножения частот одного из альтернативных признаков (p_+) на соответствующее отклонение a. Обработаем данные нашего примера указанным путем (табл. 11).

В нашей корреляционной решётке величина класса k количественного признака равна 5 дням, условное среднее $A=277{,}5$ дней. Общее число наблюдений $n=82$, частные числа наблюдений $n_+=45$ и $n_-=37$.

Находим значения pa и pa^2 для всей таблицы:

$$\Sigma pa = -41 + 47 = 6, \quad \text{откуда } \frac{\Sigma pa}{n} = \frac{6}{82} = 0{,}073;$$

$$\Sigma pa^2 = 232.$$

$p_+ \cdot a$ (для тёлочек) $= -27 + 30 = 3$, откуда

$$\Sigma \frac{p_+ \cdot a}{n_+} = \frac{3}{45} = 0{,}0666;$$

$p_- \cdot a$ (для бычков) $= -14 + 17 = 3$, откуда

$$\frac{\Sigma p_- \cdot a}{n_-} = \frac{3}{37} = 0{,}08108.$$

Отсюда длительность эмбрионального периода у телят разного пола будет следующая:

$$M_{тел} = A + k \cdot \frac{\Sigma p_+ \cdot a}{n_+} = 277{,}5 + 5 \cdot 0{,}0666 = 277{,}83 \text{ дня};$$

$$M_{быч} = A + k \cdot \frac{\Sigma p_- \cdot a}{n_-} = 277{,}5 + 5 \cdot 0{,}08108 = 277{,}9 \text{ дня}.$$

Таким образом, разница в длительности эмбрионального периода у эмбрионов разного пола незначительная и составляет 0,07 дня. Проверим это вычислением величины связи между показателем пола и длительностью эмбрионального периода. Подставим имеющиеся данные в одну из формул бисериального коэффициента (для пола тёлочек)

$$r_b = \frac{\frac{\Sigma p_+ \cdot a}{n_+} - \frac{\Sigma pa}{n}}{\sqrt{\frac{a}{n_+} - \frac{a}{n}}}.$$

Найдем значения:

$$\alpha = \Sigma pa^2 - \frac{(\Sigma pa)^2}{n} = 232 - \frac{6^2}{82} = 232 - 0{,}439 = 231{,}561.$$

$$r_b = \frac{\frac{3}{45} - \frac{6}{82}}{\sqrt{\frac{231{,}561}{45} - \frac{231{,}561}{82}}} = \frac{0{,}0666 - 0{,}0723}{\sqrt{5{,}145 - 2{,}823}} =$$

$$= -\frac{0{,}0057}{\sqrt{2{,}322}} = -\frac{0{,}0057}{1{,}523} = -0{,}0037.$$

Таблица 11

Корреляционная решётка для вычисления связи между количественным и качественным признаками

	Длительность эмбрионального развития													
	245—249	250—254	255—259	260—264	265—269	270—274	275—279 A	280—284	285—289	290—294	295—299	300—304	305—309	суммарное значение по строчкам
p_+ (телочки)	1	—	1	1	2	10	13	10	4	2	—	—	1	45
p_- (бычки)	—	1	1	—	1	3	18	9	4	—	—	—	—	37
p	1	1	2	1	3	13	31	19	8	2	—	—	1	82
a	—6	—5	—4	—3	—2	—1	0	+1	+2	+3	+4	+5	+6	—
pa	—6	—5	—8	—3	—6	—13	0	19	16	6	—	—	6	6
pa^2	36	25	32	9	12	13	0	19	32	18	—	—	36	232
$p_+ \cdot a$	—6	—	—4	—3	—4	—10	0	10	8	6	—	—	6	3
$p_- \cdot a$	—	—5	—4	—	—2	—3		9	8	—	—	—	—	3

Следовательно, бисериальный показатель связи очень мал, и поэтому связь между полом эмбриона и длительностью его эмбрионального развития практически отсутствует, что было видно и при сопоставлении средних арифметических.

6. КРИВОЛИНЕЙНЫЕ ТИПЫ СВЯЗИ

а. Корреляционное отношение η

До сих пор рассматривался прямолинейный тип связи, при котором величина и направление связи хорошо улавливались коэффициентом корреляции

В биологических явлениях часто имеет место криволинейный тип связи, о котором упоминалось на стр 72. В этих случаях коэффициент корреляции не может служить правильной мерой связи, так как он ее или приуменьшает, или не учитывает совершенно

При криволинейной связи уже нельзя пользоваться уравнением прямолинейной регрессии, в этом случае необходимо применять уравнения, в которых параметры имеют степенное значение

Недостатком коэффициента корреляции является еще то, что он не улавливает неравнозначность факторов в их взаимосвязи, когда один из них будет зависимой переменной (функцией), а другой независимой переменной (аргументом). Эта неравнозначность видна хотя бы из такого примера связь между ростом и возрастом известна, но возраст влияет на рост и определяет его значение, в то время, как обратного влияния и зависимости (роста на возраст) не может быть

В общем виде величина криволинейной связи определяется корреляционным отношением η («эта») по формуле

$$\eta_{y\,x} = \frac{\sigma(\overline{y}_x)}{\sigma_y},$$

где σ_y — среднее квадратическое отклонение для признака y; $\sigma(\overline{y}_x)$ — корень квадратный из дисперсии частных средних $\overline{y}$ около общей средней $\overline{y}$ для признака y

Так как корреляционное отношение выявляет величину связи в двух направлениях, выявляя неравнозначность воздействия x на y или y на x, то необходимо вычислить и второе значение η, в которой заменяются только соответствующие значки

Формула второго корреляционного отношения, когда зависимой величиной будет x и независимой y, выглядит следующим образом

$$\eta_{x\,y} = \frac{\sigma(\overline{x}_y)}{\sigma_x}.$$

В этих формулах $\sigma(\bar{y}_x)$ и $\sigma(\bar{x}_y)$ означают величины, полученные при извлечении квадратного корня из дисперсии σ^2.

Значения дисперсий равны следующим выражениям:

$$\sigma^2(\bar{y}_x) = \frac{1}{n}\Sigma p_x(\bar{y}_x - \bar{y})^2;$$

$$\sigma^2(\bar{x}_y) = \frac{1}{n}\Sigma p_y(\bar{x}_y - \bar{x})^2,$$

где n — общее число наблюдений;
p_x — частоты по ряду x;
p_y — частоты по ряду y;
$\bar{y}$ — средняя арифметическая по признаку y;
$\bar{x}$ — средняя арифметическая по признаку x;
$\bar{y}_x$ — средние значения y при определенном значении x;
$\bar{x}_y$ — средние значения x при определенном значении y.

Формулы вычисления корреляционного отношения для больших и малых выборок довольно сходны с той, по которой вычисляется коэффициент корреляции.

б. Вычисление корреляционного отношения для малой выборки

При малой выборке удобно вычислять η по следующим формулам:

$$\eta_{1/2} = \sqrt{\frac{\Sigma(x_1 - \bar{x}_1)^2 - \Sigma(x_1 - \bar{x}_{1.2})^2}{\Sigma(x_1 - \bar{x}_1)^2}}$$

и

$$\eta_{2/1} = \sqrt{\frac{\Sigma(x_2 - \bar{x}_2)^2 - \Sigma(x_2 - \bar{x}_{2.1})^2}{\Sigma(x_2 - \bar{x}_2)^2}}.$$

Как видно, в формулы входит сумма квадрата отклонений каждого варианта по признаку x_1 или по признаку x_2 от своей средней арифметической, т. е. от $\bar{x}_1$ или от $\bar{x}_2$. Кроме того, определяются отклонения x_1 или x_2 от частной средней $\bar{x}_{1.2}$ и $\bar{x}_{2.1}$. Разберем на примере вычисление корреляционного отношения при малом числе наблюдений ($n=10$).

Пример. Определить связь между величиной живого веса коров и удоем за лактационный период:

№ животного	Живой вес в *кг*, x_1	Удой за лактационный период в *кг*, x_2
1	500	4300
2	520	4500
3	480	4100
4	580	5000
5	600	4000
6	480	4500
7	500	4800
8	520	4000
9	600	4300
10	580	5500

Для вычисления η необходимо произвести группировку объектов по одному из признаков, например, по живому весу, на классы, объединяя объекты одного веса в один класс. При этом следует иметь в виду, что в каждом классе должно быть не меньше двух наблюдений. Если это условие отсутствует, то необходимо дополнить вариационный ряд нужным числом наблюдений по известному принципу случайной выборки.

Распределим животных выборки по показателю живого веса, выделяя следующие классы по этому признаку: 480, 500, 520, 580, 600 *кг*.

Тогда вариационный ряд будет иметь следующий вид (табл. 12).

В каждом классе по признаку x_1 оказалось по два животных. Подсчитаем средние арифметические для x_1 и x_2:

$$\bar{x}_1 = \frac{\Sigma x_1}{n} = \frac{5360}{10} = 536 \text{ кг};$$

$$\bar{x}_2 = \frac{\Sigma x_2}{n} = \frac{45000}{10} = 4500 \text{ кг}.$$

В целях применения формулы

$$\eta_{x_2 x_1} = \sqrt{\frac{\Sigma (x_2 - \bar{x}_2)^2 - \Sigma (x_2 - \bar{x}_{2\,1})^2}{\Sigma (x_2 - \bar{x}_2)^2}},$$

вычисляем $\Sigma(x_2 - \bar{x}_2)^2$ и $\Sigma(x_2 - \bar{x}_{1\,2})^2$, и полученные величины подставляем в формулу η:

$$\eta_{x_2 x_1} = \sqrt{\frac{2080000 - 500000}{2080000}} = \sqrt{0,76} = 0,8718.$$

Таблица 12

Вариационный ряд для вычисления корреляционного отношения при малом числе наблюдений между удоем и величиной живого веса

Живой вес в *кг*, x_1	Удой за лактацию в *кг*, x_2	$x_{2\cdot 1}$	$x_2 - \overline{x}_2$	$(x_2 - \overline{x}_2)^2$	$x_2 - \overline{x}_{2\cdot 1}$	$(x_2 - \overline{x}_{2\cdot 1})^2$
480	4100	$\frac{8600}{2} = 4300$	4100—4500 = —400	160000	4100—4300 = —200	40000
480	4500		4500—4500 = 0	0	4500—4300 = 200	40000
500	4300	$\frac{9100}{2} = 4550$	4300—4500 = —200	40000	4300—4550 = 250	62500
500	4800		4800—4500 = 300	90000	4800—4550 = —250	62500
520	4500	$\frac{8500}{2} = 4250$	4500—4500 = 0	0	4500—4250 = 250	62500
520	4000		4000—4500 = —500	250000	4000—4250 = —250	62500
580	5000	$\frac{10500}{2} = 5250$	5000—4500 = 500	250000	5000—5250 = —250	62500
580	5500		5500—4500 = 1000	1000000	5500—5250 = 250	62500
600	4000	$\frac{8300}{2} = 4150$	4000—4500 = —500	250000	4000—4150 = —150	22500
600	4300		4300—4500 = —200	40000	4300—4150 = 150	22500
			$\Sigma(x_2 - \overline{x}_2) = 0$	$\Sigma(x_2 - \overline{x}_2)^2 = 2080000$	$\Sigma(x_2 - \overline{x}_{2\cdot 1}) = 0$	$\Sigma(x_2 - \overline{x}_{2\cdot 1})^2 = 500000$

Из полученного значения корреляционного отношения первого признака по второму следует, что удой в большой степени зависит от величины живого веса.

Вычисления корреляционного отношения можно вести и по таким формулам (при однозначных значениях вариантов удобно возвести их в квадрат):

$$\eta_{2/1} = \sqrt{\frac{C_{2\cdot 1}}{C_2}} \quad \text{и} \quad \eta_{1/2} = \sqrt{\frac{C_{1\cdot 2}}{C_1}},$$

где $$C_{2\cdot 1} = \Sigma h - H \quad \text{и} \quad C_2 = \Sigma x^2 - H.$$

В последних выражениях значения $C_{2\cdot 1}$ и $C_{1\cdot 2}$ представляют собой показатели частной дисперсии, а значения C_2 и C_1 — общей дисперсии, что будет подробнее объяснено в главе о дисперсионном анализе.

Известно, что:

$$\Sigma h = \Sigma\left\{\frac{(\Sigma x_{2\cdot 1})^2}{n_1}\right\} \quad \text{и} \quad H = \frac{(\Sigma x_2)^2}{n}.$$

Разберем это на примере зависимости урожайности плодовых деревьев от их возраста.

Пример. Имеются следующие данные о возрасте и урожайности плодовых деревьев, которые сгруппированы по возрасту в классы (см. стр. 102).

Требуется определить величину связи между возрастом и урожайностью плодовых деревьев путем вычисления корреляционного отношения.

Произведем группировку растений в классы по показателю возраста:

Подставим полученные данные в выражения h и H:

$$\Sigma h = \Sigma\left\{\frac{(\Sigma x_{2\cdot 1})^2}{n_1}\right\} = 104{,}1; \; H = \frac{(\Sigma x_2)^2}{n} = \frac{30^2}{10} = 90,$$

откуда:

$$C_{2\cdot 1} = \Sigma h - H = 104{,}1 - 90 = 14{,}1;$$

$$C_2 = \Sigma x_2^2 - H = 110 - 90 = 20.$$

Находим значение корреляционного отношения:

$$\eta_{2/1} = \sqrt{\frac{C_{2\cdot 1}}{C_2}} = \sqrt{\frac{14{,}1}{20}} = \sqrt{0{,}705} = 0{,}83;$$

№ растения	Возраст в годах, x_1	Урожайность в ц, x_2	$x_{2\cdot 1}$	n_1	x_2^2	$(\Sigma x_{2\cdot 1})^2$	$h = \frac{(\Sigma x_{2\cdot 1})^2}{n_1}$
1	3	1			1		
2	3	2	1+2+1=4	3	4	16	16 : 3=5,3
3	3	1			1		
4	4	2	2+4=6	2	4	36	36 : 2=18
5	4	4			16		
6	5	4	4+5=9	2	16	81	81 : 2=40,5
7	5	5			25		
8	6	5			25		
9	6	3	5+3+3=11	3	9	121	121 : 3=40,3
10	6	3			9		
	$\Sigma x_1=45$	$\Sigma x_2=30$	$\Sigma x_{2\cdot 1}=30$	$\Sigma n_1=\Sigma n=10$	$\Sigma x_2^2=110$	—	$\Sigma h=104,1$

следовательно, величина урожайности x_2 существенно обусловлена влиянием возраста.

Как уже отмечалось, C_2 выражает общую дисперсию варьирующего признака. Под общей дисперсией (изменчивостью) признака подразумевается та изменчивость, которая формируется под воздействием многих и разнообразных факторов. В нашем примере на урожайность могли влиять возраст, удобрение, условия климата, сорт и другие факторы.

Под частной дисперсией признака $C_{2.1}$ подразумевается доля изменчивости, вызываемая воздействием фактора x_2, т. е. возрастом.

Таким образом, корреляционное отношение, выявляя общую связь признака с воздействующим фактором, связано с уровнем общей и частной дисперсий.

Если вычислить величину связи в нашем примере в обратном направлении, т. е. определить связь возраста с урожайностью, то, по данным разбираемого примера, получается $\eta_{1/2}=0,94$. *Для получения этого значения требуется распределить* объекты в классы по урожайности, и вариационный ряд будет иметь следующий вид (см. стр. 104).

Следовательно:

$$\Sigma h = 215,5;\ H = \frac{(\Sigma x_1)^2}{n} = \frac{45^2}{10} = 202,5.$$

Отсюда:

$$C_{1\cdot 2} = \Sigma h - H = 215,5 - 202,5 = 13,0;$$

$$C_2 = \Sigma x_1^2 - H = 217 - 202,5 = 14,5.$$

Вычислим второе значение корреляционного отношения:

$$\eta_{1/2} = \sqrt{\frac{13,0}{14,5}} = \sqrt{0,9} = 0,9487.$$

Значение η показывает, что зависимость возраста от урожайности велика. Несмотря на биологический абсурд такой постановки вопроса, когда возраст как бы зависит от урожайности, в математическом смысле абсурда нет. Можно также *найти и смысловое значение «зависимости возраста от урожайности»*, а именно, по величине урожайности можно судить о возможном возрасте.

№ растения	Урожайность в ц, x_2	Возраст в годах, x_1	x_1^2	n_2	$\Sigma x_{1\cdot 2}$	$(\Sigma x_{1\cdot 2})^2$	$h = \frac{(\Sigma x_{1\cdot 2})^2}{n_2}$
1 3	1 1	3 3	9 9	2	$3+3=6$	36	$36:2=18$
2 4	2 2	3 4	9 16	2	$3+4=7$	49	$49:2=24{,}5$
9 10	3 3	6 6	36 36	2	$6+6=12$	144	$144:2=72$
5 6	4 4	4 5	16 25	2	$4+5=9$	81	$81:2=40{,}5$
7 8	5 5	5 6	25 36	2	$5+6=11$	121	$121:2=60{,}5$
	$\Sigma x_2=30$	$\Sigma x_1=45$	$\Sigma x_1^2=217$	$\Sigma n_2=\Sigma n=10$	$\Sigma x_{1\cdot 2}=45$	—	$\Sigma h=215{,}5$

Аналогичная постановка вопроса и ответ на него часто используются в жизни. Так, например, школьник-первоклассник дает представление о том, что возраст такого ребенка около 7 лет; такое же суждение о возрасте можно сделать и по росту ребенка. В животноводческой практике широко используется суждение о возрасте животного по стертости зубной системы.

Из этих примеров видно, что корреляционное отношение хорошо выявляет неравнозначность связи между коррелирующими признаками. Это свойство корреляционного отношения имеет существенное значение при составлении биологических прогнозов, так как по величине связи между независимым фактором воздействия и зависимым признаком биологического характера с помощью уравнения криволинейной регрессии можно определить значение зависимого фактора при том или ином значении независимого признака. Так, например, по показателям температуры или влажности можно делать прогноз скорости размножения и роста численности вредителей, по уровню рациона — планировать привесы или удой, по дозе облучения — ожидать определенную степень лучевого поражения.

в. Вычисление корреляционного отношения для большой выборки

При вычислении корреляционного отношения для большой выборки производится составление корреляционной решётки таким же методом, как это было показано при вычислении коэффициента корреляции (см. стр. 82). При этом можно воспользоваться следующими рабочими формулами, для которых необходимые данные получаются при обработке корреляционной решётки методом произведений (или методом сумм) и условного отклонения:

$$\eta_{2/1} = \sqrt{\frac{\sum\left[\frac{(\Sigma p_{1\cdot 2}\cdot a_2)^2}{p_1}\right] - \frac{(\Sigma p_2\cdot a_2)^2}{n}}{\alpha_2}},$$

где

$$\alpha_2 = \Sigma p_2\cdot a_2^2 - \frac{(\Sigma p_2\cdot a_2)^2}{n};$$

$$\eta_{1/2} = \sqrt{\frac{\sum\left[\frac{(\Sigma p_{1\cdot 2}\cdot a_1)^2}{p_2}\right] - \frac{(\Sigma p_1\cdot a_1)^2}{n}}{\alpha_1}},$$

где

$$\alpha_1 = \Sigma p_1 \; a_1^2 - \frac{(\Sigma p_1 \; a_1)^2}{n}$$

В приведенных формулах

p_1 — частоты ряда по первому признаку,

p_2 — частоты ряда по второму признаку,

n — число наблюдений в выборке,

a_1 и a_2 — условные отклонения от условных средних A_1 и A_2 каждого признака

Выражения, используемые в формулах α_1 и α_2, известны из вычислений, которые рассматривались при определении коэффициента корреляции на стр 83, где определялись суммы по каждому признаку для рядов pa и pa^2, входящих и в формулу корреляционного отношения Неизвестным остается способ вычисления выражений

$$\sum \frac{(\Sigma p_{1\;2} \; a_2)^2}{p_1} \quad \text{и} \quad \sum \frac{(\Sigma p_{1\;2} \; a_1)^2}{p_2} ,$$

что будет показано ниже

В качестве примера на вычисление корреляционного отношения для большой выборки проведем обработку корреляционной решётки, составленной из показателей жирномолочности коров джерсейской породы за лактационный период, т е из показателей среднелактационного процента жира и из показателей календарных сроков отёла этих коров Требуется определить, будут ли иметь место связь и зависимость жирномолочности за лактационный период от календарных сроков отёла коров

На стр 107 приводится корреляционная решетка (табл 13).

Проводим выделение классов с условными средними A_1 и A_2 Составляем ряды условных отклонений по обоим признакам a_1 и a_2, вычисляем показатели p_1a_1 для ряда, изображающего процент жира, и p_2a_2 для ряда, изображающего сроки отёла Вычисляем ряды $p_1a_1^2$ и $p_2a_2^2$ Подсчитываем алгебраическую сумму по показателям pa и pa^2 для обоих признаков и получаем, что $\Sigma p_1a_1=60$, $\Sigma p_1a_1^2 = 1052$, $\Sigma p_2a_2=0$, $\Sigma p_2a_2^2 = 3650$

В формулу $\eta\;^2/_1$ входит выражение $\Sigma \; \frac{(\Sigma p_{1\;2} \; a_2)^2}{p_1}$

Для получения этого выражения требуется умножить частоты, стоящие в каждой строчке $(p_{1\;2})$, на соответствующее

Таблица 13

Корреляционная решётка связи жиромолочности с календарным сроком отёла

Средний процент жира за лактационный период (1)	Календарный месяц отёла (2): Январь	Февраль	Март	Апрель	Май	Июнь	Июль	Август A_1	Сентябрь	Октябрь	Ноябрь	Декабрь	p_1	a_1	$p\ a_1$	$p_1\ a_1^2$	$p_{1\ 2}\ a_1$	$(p_{1\ 2}\ a)^2$	$h=(p_{1\ 2}\ a)^2:p_1$
4,8—4,9	—	—	—	—	—	—	—	2	—	—	—	—	2	−5	−10	50	0	0	0
5,0—5,1	—	—	—	—	—	—	3	2	—	—	—	2	7	−4	−28	112	+5	25	3,57
5,2—5,3	1	1	—	—	1	—	—	2	3	2	1	—	11	−3	−33	99	−6	36	3,27
5,4—5,5	1	5	1	1	2	3	5	4	2	6	4	1	35	−2	−70	140	−33	1089	31,1
5,6—5,7	4	4	3	2	7	4	3	5	9	12	7	7	67	−1	−67	67	−25	625	9,33
A_1 5,8—5,9	7	5	4	2	1	5	3	6	9	12	14	9	77	0	0	0	−12	144	1,87
6,0—6,1	8	5	2	—	2	2	2	12	13	24	9	4	83	1	83	83	−4	16	0,19
6,2—6,3	2	—	—	1	1	—	5	6	11	12	11	2	51	2	102	204	+50	2500	49,0
6,4—6,5	1	—	—	—	—	—	1	1	4	4	4	—	15	3	45	135	+16	256	17,06
6,6—6,7	—	—	—	—	—	—	—	2	2	3	—	—	7	4	28	112	+8	64	9,1
6,8—6,9	—	—	—	—	—	—	—	1	1	—	—	—	2	5	10	50	+1	1	0,5
p_2	24	20	10	6	14	14	22	43	54	75	50	25	357	—	60	1052	—	—	124,99
a_2	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	—	—	—	—	—	—	—
$p_2 \cdot a_2$	−168	−120	−50	−24	−42	−28	−22	0	54	150	150	100	−454 +454 0	—	—	—	—	—	—
$p_2 \cdot a_2^2$	1176	720	250	96	126	56	22	0	54	300	450	400	3650	—	—	—	—	—	—
$p_{1\cdot 2} \cdot a_1$	+6	−12	−3	−2	−10	−8	−10	+3	+38	+42	+25	−9	—	—	—	—	—	—	—
$(p_{1\ 2} \cdot a_1)^2$	36	144	9	4	100	64	100	9	1444	1764	625	81	—	—	—	—	—	-	—
$h_2 = (p_{1\ 2} \cdot a_1)^2 : p_2$	1,5	7,2	0,9	0,66	7,14	4,57	4,54	0,20	26,74	23,52	12,5	3,24	92,71	—	—	—	—	—	—
$M_{1\cdot 2} = \frac{\Sigma p_{1\ 2} \cdot a_1}{p_2} k_1 + A_1$	5,95	5,78	5,84	5,83	5,76	5,79	5,81	5,91	6,04	6,01	6,0	5,83	—	—	—	—	—	—	—

отклонение a_2. Делаем это по первой строчке и получаем: $p_{1\,2} \cdot a_2 = 2 \cdot 0 = 0$; записываем это в соответствующую графу.

То же самое делаем по второй строчке:

$$p_{1\,2} \cdot a_2 = 3 \cdot (-1) + 2 \cdot (0) + 2 \cdot (4) = +5.$$

Продолжим по этому же принципу расчет показателей $p_{1\,2} \cdot a_2$ для остальных девяти строчек:

3-я строчка: $1 \cdot (-7) + 1 \cdot (-6) + 1 \cdot (-3) + 2 \cdot (0) + 3 \cdot (+1) + 2 \cdot (+2) + 1 \cdot (+3) = -6$;

4-я строчка: $1 \cdot (-7) + 5 \cdot (-6) + 1 \cdot (-5) + 1 \cdot (-4) + 2 \cdot (-3) + 3 \cdot (-2) + 5 \cdot (-1) + 4 \cdot (0) + 2 \cdot (+1) + 6 \cdot (+2) + 4 \cdot (+3) + 1 \cdot (+4) = -33$;

5-я строчка: $4 \cdot (-7) + 4 \cdot (-6) + 3 \cdot (-5) + 2 \cdot (-4) + 7 \cdot (-3) + 4 \cdot (-2) + 3 \cdot (-1) + 5 \cdot (0) + 9 \cdot (+1) + 12 \cdot (+2) + 7 \cdot (+3) + 7 \cdot (+4) = -25$;

6-я строчка: $7 \cdot (-7) + 5 \cdot (-6) + 4 \cdot (-5) + 2 \cdot (-4) + 1 \cdot (-3) + 5 \cdot (-2) + 3 \cdot (-1) + 6 \cdot (0) + 9 \cdot (+1) + 12 \cdot (+2) + 14 \cdot (+3) + 9 \cdot (+4) = -12$;

7-я строчка: $8 \cdot (-7) + 5 \cdot (-6) + 2 \cdot (-5) + 2 \cdot (-3) + 2 (-2) + 2 \cdot (-1) + 12 \cdot (0) + 13 \cdot (+1) + 24 \cdot (+2) + 9 \cdot (+3) + 4 \cdot (+4) = -4$;

8-я строчка: $2 \cdot (-7) + 1 \cdot (-4) + 1 \cdot (-3) + 5 \cdot (-1) + 6 \cdot (0) + 11 \cdot (+1) + 12 \cdot (+2) + 11 \cdot (+3) + 2 \cdot (+4) = +50$;

9-я строчка: $1 \cdot (-7) + 1 \cdot (-1) + 1 \cdot (0) + 4 \cdot (+1) + 4 \cdot (+2) + 4 \cdot (+3) = +16$;

10-я строчка: $2 \cdot (0) + 2 \cdot (+1) + 3 \cdot (+2) = +8$;

11-я строчка: $1 \cdot (0) + 1 \cdot (+1) = +1$.

Таким образом, для каждой строчки значения $p_{1\,2} \cdot a_2$ получены следующие: 0, +5, −6, −33, −25, −12, −4, +50, +16, +8, +1. Эти данные и записываются в корреляционной таблице в столбик под соответствующим заголовком. Последующая обработка заключается в возведении этих данных в квадрат, что записывается в следующем столбике с заголовком $(p_{1\,2} \cdot a_2)^2$. Последнее действие заключается в делении полученных значений $(p_{1\,2} \cdot a_2)^2$ на величину p_1, т. е. на частоты вариационного ряда первого признака. Проделаем эти вычисления и просуммируем величины $(p_{1 \cdot 2} \cdot a_2)^2 : p_1$:

$$\Sigma h_1 = \Sigma\{(p_{1\,2}\cdot a_2)^2 : p_1\} = \frac{0}{2} + \frac{25}{7} + \frac{36}{11} + \frac{1089}{35} + \frac{625}{67} + \frac{144}{77} + \frac{16}{83} + \frac{2500}{51} + \frac{256}{15} + \frac{64}{7} + \frac{1}{2} = 0 + 3{,}57 + 3{,}27 + 31{,}1 + 9{,}33 + 1{,}87 + 0{,}19 + 49{,}0 + 17{,}06 + 9{,}1 + 0{,}5 = 124{,}99\cdot$$

Аналогичные расчеты делаются для вычисления величины $(p_{1\,2}\cdot a_1)^2 . p_2$, входящей в формулу $\eta_{1/2}$.

Находим произведения $p_{1\,2}\cdot a_1$ для каждого столбца, умножая частоты каждой клеточки в первом столбце на соответствующее значение a_1. Получаем следующие значения $p_{1\,2}\cdot a_1$:

1-й столбец: $1\cdot(-3) + 1\cdot(-2) + 4\cdot(-1) + 7\cdot(0) + 8\cdot(+1) + 2\cdot(+2) + 1\cdot(+3) = +6$;
2-й столбец: $1\cdot(-3) + 5\cdot(-2) + 4\cdot(-1) + 5\cdot(0) + 5\cdot(+1) = -12$;
3-й столбец: $1\cdot(-2) + 3\cdot(-1) + 4\cdot(0) + 2\cdot(+1) = -3$;
4-й столбец: $1\cdot(-2) + 2\cdot(-1) + 2\cdot(0) + 1\cdot(+2) = -2$;
5-й столбец: $1\cdot(-3) + 2\cdot(-2) + 7\cdot(-1) + 1\cdot(0) + 2\cdot(+1) + 1\cdot(+2) = -10$;
6-й столбец: $3\cdot(-2) + 4\cdot(-1) + 5\cdot(0) + 2\cdot(+1) = -8$;
7-й столбец: $3\cdot(-4) + 5\cdot(-2) + 3\cdot(-1) + 3\cdot(0) + 2\cdot(+1) + 5\cdot(+2) + 1\cdot(+3) = -10$;
8-й столбец: $2\cdot(-5) + 2\cdot(-4) + 2\cdot(-3) + 4\cdot(-2) + 5\cdot(-1) + 6\cdot(0) + 12\cdot(+1) + 6\cdot(+2) + 1\cdot(+3) + 2\cdot(+4) + 1\cdot(+5) = +3$;
9-й столбец: $3\cdot(-3) + 2\cdot(-2) + 9\cdot(-1) + 9\cdot(0) + 13\cdot(+1) + 11\cdot(+2) + 4\cdot(+3) + 2\cdot(+4) + 1\cdot(+5) = +38$;
10-й столбец: $2\cdot(-3) + 6\cdot(-2)\,12\cdot(-1) + 12\cdot(0) + 24\cdot(+1) + 12\cdot(+2) + 4\cdot(+3) + 3\cdot(+4) = +42$;
11-й столбец: $1\cdot(-3) + 4\,(-2) + 7\cdot(-1) + 14\cdot(0) + 9\cdot(+1) + 11\cdot(+2) + 4\cdot(+3) = +25$;
12-й столбец: $2\cdot(-4) + 1\cdot(-2) + 7\,(-1) + 9\cdot(0) + 4\cdot(1) + 2\cdot(2) = -9$.

Далее возводим каждое значение $(p_{1\,2}\cdot a_1)$ в квадрат и *получаем* для каждого столбца строчку значений $(p_{1\,2}\cdot a_1)^2$, которая запишется так: 36; 144; 9; 4; 100; 64; 100; 9; 1444; 1764; 625; 81.

Составляем последнюю строчку из значений $(p_{1\,2}\cdot a_1)^2 : p_2$, т. е. делим каждое значение $(p_{1\,2}\cdot a_1)^2$ на соответствующее значение частот ряда второго признака (p_2) и суммируем:

$$\Sigma h_2 = \Sigma \{(p_{1\cdot 2} \cdot a_1)^2 : p_2\} = \frac{36}{24} + \frac{144}{20} + \frac{9}{10} + \frac{4}{6} + \frac{100}{14} + \frac{64}{14} + \frac{100}{22} + \frac{9}{43} + \frac{1444}{54} + \frac{1764}{75} + \frac{625}{50} + \frac{81}{25} =$$

$$= 1,5 + 7,2 + 0,9 + 0,66 + 7,14 + 4,57 + 4,54 + 0,20 + 26,74 + 23,52 + 12,5 + 3,24 = 92,71.$$

Обработка данных корреляционной решётки из табл. 13

	Январь	Февраль	Март	Апрель	Май
$p_{1\cdot 2} \cdot a_1$	6	−12	−3	−2	−10
$\frac{p_{1\cdot 2} \cdot a_1}{p_2}$	$\frac{6}{24}$	$\frac{-12}{20}$	$\frac{-3}{10}$	$\frac{-2}{6}$	$\frac{-10}{14}$
$\frac{p_{1\cdot 2} \cdot a_1}{p_2}$	0,25	−0,60	−0,30	−0,33	−0,71
$\frac{p_{1\cdot 2} \cdot a_1}{p_2} \cdot k_1^*$	0,05	−0,12	−0,06	−0,066	−0,142
$\frac{p_{1\cdot 2} \cdot a \cdot k}{p_2} + A_1^*$	5,95	5,78	5,84	5,83	5,76

$k_1^* = 0,2\%$; $A_1^* = 5,90\%$.

Вычисляем значения α_1 и α_2, входящих в знаменатель формулы корреляционного отношения:

$$\sigma_1 = \Sigma p_1 \cdot a_1^2 - \frac{(\Sigma p_1 \cdot a_1)^2}{n} = 1052 - \frac{60^2}{357} = 1052 - 10,08 = 1041,92;$$

$$\sigma_2 = \Sigma p_2 \cdot a_2^2 = \frac{(\Sigma p_2 \cdot a_2)^2}{n} = 3650 - \frac{0}{357} = 3650.$$

Теперь имеются все данные для вычисления обоих значений корреляционного отношения.

Подставляем соответствующие значения в формулу η:

$$\eta_{2/1} = \sqrt{\frac{\sum \frac{(p_{1\cdot 2}\cdot a_2)^2}{p_1} - \frac{(\Sigma p_2\cdot a_2)^2}{n}}{a_2}} =$$

$$= \sqrt{\frac{124,99 - \frac{0}{357}}{3650}} = \sqrt{\frac{124,99}{3650}} = \sqrt{0,0342} = 0,1849;$$

Таблица 14

для вычисления показателей эмпирической линии регрессии

Июнь	Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
−8	−10	3	38	42	25	−9
$\frac{-8}{14}$	$\frac{-10}{22}$	$\frac{3}{43}$	$\frac{38}{54}$	$\frac{42}{75}$	$\frac{25}{50}$	$\frac{-9}{25}$
−0,57	−0,45	0,07	0,7	0,56	0,50	−0,36
−0,114	−0,09	0,014	0,14	0,112	0,100	−0,072
5,79	5,81	5,91	6,04	6,01	6,0	5,83

следовательно, связь календарного срока отёла с жирномолочностью незначительна.

Если же вычислить корреляционное отношение зависимости жирномолочности коров от календарного срока их отёла $\eta_{1/2}$, то связь между этими признаками оказывается большой, а именно:

$$\eta_{1/2} = \sqrt{\frac{\sum \frac{(p_{1\cdot 2}\cdot a_1)^2}{p_2} - \frac{(\Sigma p_1\cdot a_1)^2}{n}}{a_1}} =$$

$$= \sqrt{\frac{92,71 - \frac{60^2}{357}}{1041,92}} = \sqrt{\frac{92,71 - 10,08}{1041,92}} = \sqrt{\frac{82,63}{1041,92}} =$$

$$= \sqrt{0,793} = 0,89.$$

Значит, календарный срок отёла в большой мере влияет на жирномолочность коров.

Для подтверждения этой зависимости можно построить эмпирическую линию регрессии. Для этого в корреляционной решётке следует вычислить *частные средние значения каждого признака при определенном значении другого признака*, т. е. следует найти частные средние арифметические $M_{1 \cdot 2}$ и $M_{2 \cdot 1}$

Для этого можно воспользоваться следующей рабочей формулой:

$$M_{1\,2} = \frac{\Sigma p_{1\,2} \cdot a_1}{p_2} \cdot k_1 + A_1,$$

где $p_{1\,2}$ — частоты в клетках корреляционной решётки при определенном значении второго признака;

a_1 — условное отклонение по первому признаку;

k_1 — величина классового промежутка по первому признаку;

A_1 — условная средняя первого признака $\left(A_1 = 5,8 + \frac{k_1}{2} = = 5,8 + 0,1 = 5,9\right)$.

Значение $p_{1\,2} \cdot a_1$ вычислено и представлено четвертой строчкой снизу в корреляционной таблице.

Расчет этой величины следующий (см. табл. 14).

В табл. 13 на стр. 107 приведены в последней строке вычисленные значения величины $M_{1\,2}$.

Из приведенных расчетов частных средних процента жира в молоке коров для каждого календарного месяца отёла видно, что имеет место определенная закономерность изменений процента жира в молоке коров в зависимости от сезона отёла.

Так, при весенне-летних сроках отёла процент жира за лактационный период оказался меньше, чем при осенне-зимних сроках. Эта зависимость особенно хорошо выявляется на графике (рис. 19), изображающем эмпирическую линию регрессии жирномолочности в связи с календарным месяцем отёла.

Линия регрессии ясно указывает на *криволинейный тип* связи, которая могла быть выявлена только корреляционным отношением.

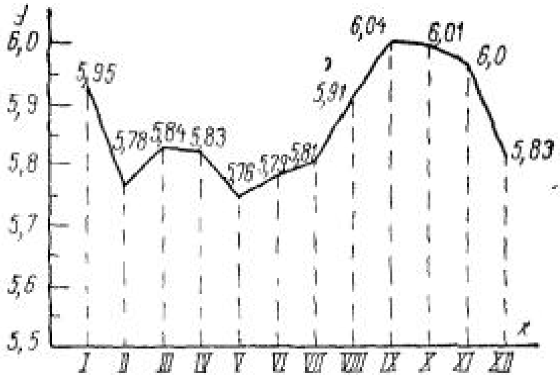

Рис. 19

О степени криволинейности можно судить по величине показателя криволинейности L:

$$L = \eta^2 - r^2,$$

где η^2 — квадрат корреляционного отношения; r^2 — квадрат коэффициента корреляции. Если связь прямолинейна, то $L=0$. Если связь криволинейна, то $L>0$.

ГЛАВА VII

СТАТИСТИЧЕСКИЕ ОШИБКИ

(мера точности статистических характеристик)

1. ТИПЫ ОШИБОК И МЕТОДЫ ИХ УМЕНЬШЕНИЯ

Статистический метод позволяет получать характеристики средних *величин, величину и направление связи* между признаками и величину изменчивости признаков. Эти вопросы ранее были разобраны нами. Тогда же отмечалось, что при вычислении этих показателей можно пользоваться выборочной совокупностью, которая составляет только часть генеральной. Методом случайного отбора членов выборки достигается правильное отражение выборкой закономерностей генеральной совокупности.

Принцип же суждения о целом по его части служит источником *ошибок статистического* характера.

Величина статистической ошибки зависит от того, как велико число наблюдений, попавших из генеральной совокупности в случайную выборку, причем, чем больше число наблюдений, тем меньше величина статистической ошибки и тем ближе показатели, полученные на выборочном материале, приближаются к показателям генеральной совокупности.

На величину статистической ошибки влияет степень изменчивости изучаемых признаков. Чем больше изменчив признак, тем больше величина статистической ошибки. Если бы все особи генеральной совокупности были одинаковы, т. е. не имели бы изменчивости, то в таком случае выборка, составленная даже из одного наблюдения, характеризовала бы свойства любой особи, входящей в генеральную совокупность. В массовых же явлениях приходится сталкиваться с вариабильностью признаков даже в пределах узких границ однородной совокупности. Это свойство объектов и будет увеличивать статистическую ошибку.

При статистическом анализе необходимо стремиться к уменьшению статистических ошибок, т. е. необходимо влиять на те показатели, от которых они зависят.

Но так как влиять на уменьшение изменчивости признаков обследованных объектов нельзя, то для уменьшения величины ошибки имеется только один путь: увеличить число наблюдений, образующих случайную выборку.

Поэтому в научной работе очень важно еще до постановки опыта или до проведения обследования правильно определить, какое же число наблюдений нужно ввести в случайную выборку, чтобы получить допустимую величину статистических ошибок.

Для этого служат три способа:

а. Если известна степень изменчивости признака, выраженная в процентах коэффициентом изменчивости $\left(C = \frac{\sigma}{M} \cdot 100\right)$, то можно пользоваться номограммой достаточно больших чисел.

Номограмма строится из трех вертикальных линий: левая вертикальная изображает коэффициент изменчивости признака C в процентах; крайняя правая указывает на требуемую точность вычисления и величину ошибки E, выраженную в процентах (рис. 20); на третьей линии, находящейся между двумя крайними, нанесены отрезки, указывающие на число наблюдений n, которое необходимо ввести в выборку при заданной ве-

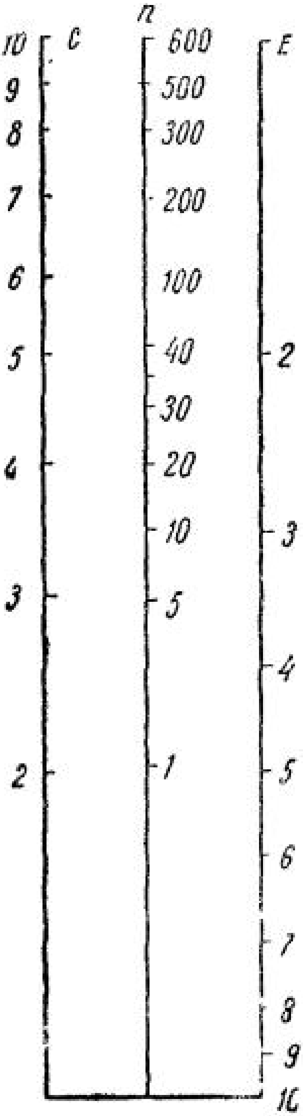

Рис 20

личине ошибки E и определенном коэффициенте изменчивости C.

Отрезки на всех линиях нанесены в определенном соотношении между собой. Пользоваться номограммой нужно следующим образом: если признак имеет коэффициент изменчивости $C=6\%$ и ставится цель получить величину ошибки $E=5\%$, то, соединяя указанные точки на крайних линиях линейкой или ниткой, на средней линии в месте пересечения получаем точку, соответствующую числу наблюдений, необходимых для включения в выборку. Так, в данном примере число наблюдений 15.

б. Если же коэффициент изменчивости признака неизвестен, то необходимое число наблюдений для включения в выборку определяется по таблице достаточно больших чисел, обеспечивающих получение искомого числа наблюдений при любом коэффициенте изменчивости признака.

Эта таблица выглядит так:

Таблица 15

Таблица достаточно больших чисел

E (величина ошибки в %)	P (вероятность)			
	0,95	0,99	0,995	0,999
5	384	663	787	1082
4	600	1036	1231	1691
3	1067	1843	2188	3007
2	2400	4146	4924	6767
1	9603	16587	19699	27069

По этой таблице искомое число n получается больше чем по номограмме, так как неизвестен коэффициент изменчивости признака.

Из предыдущих пояснений (см. стр. 11) известно, что чем ближе вероятность события P к единице, тем больше шансов за то, что это событие будет осуществляться.

Если $P=0{,}95$, то это означает, что из 100 случаев в 95 это событие будет иметь место; если $P=0{,}999$, то из 1000 в 999 случаях это событие осуществится и только в одном случае его не будет.

Применяя показатель вероятности к статистическим ошибкам, из таблицы можно заметить, что чем меньше требуется допустить ошибку, тем большее число наблюдений нужно иметь, и чем с большей вероятностью желательно получить соответствие между выборочными данными и показателями генеральной совокупности, тем еще больше требуется иметь наблюдений в выборке.

Если между выборочной и генеральной совокупностями допускается расхождение, равное Δ, то оно не должно превышать величины тройной статистической ошибки m, т. е. $\Delta \leqslant 3m$. Величину Δ можно выразить в долях среднего квадратического отклонения $\Delta = \sigma \cdot d = 3m$. Отсюда $d = \frac{\Delta}{\sigma}$, что выражает допустимое расхождение средних данных выборки от средних, характеризующих генеральную совокупность. При этом расхождение d выражено в долях сигмы.

Если известен объем (т. е. численность) генеральной совокупности N и задана допустимая точность расхождения d, то необходимое число наблюдений, которое требуется включить в выборку, можно вычислить по следующей формуле-

$$n = \frac{N}{N\left(\frac{d}{3}\right)^2 + 1},$$

где n — искомое число наблюдений для выборки;

N — численность генеральной совокупности;

$d = \frac{\Delta}{\sigma}$ — допускаемое расхождение, выраженное в долях сигмы, или заданная точность величин;

Δ — допускаемое расхождение между выборочной и генеральной средними;

σ — среднее квадратическое отклонение.

Если $N = \infty$, то искомое значение $n = \frac{9}{d^2}$ ($9 = t^2 = 3^2$, что означает критерий достоверности при $P = 0{,}997$).

Пример. На контрольной и подопытной делянках имеется по 10000 растений картофеля.

Сколько требуется взять растений в выборку, чтобы получить достоверное соответствие между выборочной и генеральной средними, если при изучении веса клубней $\sigma = 600$ *г*, а расхождение $\overline{M}$ (генеральной) с M (выборочной) должно быть не более 100 *г*?

Находим:

$$d = \frac{\Delta}{\sigma} = \frac{100}{600} = \frac{1}{6}; \quad d^2 = \left(\frac{1}{6}\right)^2 = \frac{1}{36} = 0{,}028,$$

откуда:

$$n = \frac{10000}{10000 \cdot \left(\frac{0,028}{9}\right) + 1} = 322,$$

т. е. 322 растения должны быть обследованы и включены в выборку.

Если требования к уровню вероятности снизить и взять его *не при* $P=0{,}997$ (когда $t = 3$), *а при* $P=0{,}95$ (когда $t=2$), то число необходимых наблюдений будет снижаться.

Пример. Предположим, требуется определить количество мышей, которых нужно взять в опыт для изучения влияния облучения на плодовитость. Изменчивость плодовитости выражается средним квадратическим отклонением $\sigma = 2$ головам. Допускаемое расхождение $\Delta=1$ голове. Отсюда $d=\frac{1}{2}=$ $=0{,}5$. Генеральная совокупность имеет численность в 1000 голов. Тогда численность выборки должна быть равна.

$$n = \frac{N}{N \cdot \left(\frac{d}{3}\right)^2 + 1} = \frac{1000}{1000 \cdot \left(\frac{0,5}{3}\right)^2 + 1} = \frac{1000}{1000 \cdot \frac{0,25}{9} + 1} =$$

$$= \frac{1000}{\frac{250}{9} + 1} = \frac{1000}{27,7 + 1} = \frac{1000}{28,7} \approx 32 \text{ головы.}$$

Таким образом, при данной изменчивости признака и заданной точности опыта требуется ввести в опыт не менее 32 самок мышей для изучения их плодовитости под влиянием облучения, что обеспечит получение статистически достоверных данных.

Если численность выборки устанавливается для членов совокупности по качественным признакам, то все расчеты остаются аналогичными вышеприведенным, меняется только способ определения d, когда вместо σ берется выражение $\sqrt{pq}$, где p и q являются предполагаемыми долями встречаемости данного признака у членов генеральной совокупности. Например, доля особей у помесей второго поколения, имеющих доминантный признак, по правилу Менделя, равна 0,75, а доля особей с рецессивным признаком будет 0,25. Если доли встречаемости признаков в генеральной совокупности неизвестны, то условно берется, что $p=q=0{,}5$, откуда

$$d = \frac{\Delta}{\sqrt{0,5 \cdot 0,5}} = \frac{\Delta}{0,5} = 2\Delta.$$

Кроме статистических ошибок в работе могут встречаться и другие типы ошибок.

Так, осуществляя измерение какого-либо показателя, может допускаться неточность в работе за счет недостаточной точности используемого прибора. Например, измеряя высоту растения, можно пользоваться линейкой, у которой имеются деления, равноценные 0,1 *см*, при этом измерение пойдет с точностью до 0,1 *см*, но можно пользоваться или более точным, или более грубым мерным инструментом, и тогда точность получения показателей высоты растений будет или больше, или меньше. Поэтому при работе необходимо знать точность прибора и вводить соответствующую поправку на техническую возможность точности прибора.

Часто к приборам прилагается *заводской паспорт*, в котором сравниваются показатели данного прибора с показателями прибора, служащего эталоном и хранящимся в палате мер и весов. Зная точность измерения данного прибора, следует вводить в измерения, полученные с его помощью, необходимые поправки.

Например, на паспорте к гемометру Сали указано, что он дает отсчет с ошибкой $+0,5$ единиц Сали. Следовательно, если сделать с его помощью отсчет и получить значение показателя гемоглобина, равное 85 единицам, то, учтя ошибку прибора, истинный показатель будет равен 85,5 единицам.

Могут быть в работе ошибки, вызванные описками, просчетами, небрежностью. Эти ошибки необходимо своевременно выявлять и устранять. Неучтенность этих ошибок может дать совершенно искаженные показатели и привести к неверным выводам.

Большую опасность для правильности научных выводов несут методические ошибки, вызванные неправильной схемой опыта, неправильным выбором объектов исследования, субъективным отношением к их отбору и т. п. Предупреждение этих ошибок может быть осуществлено продуманной методикой и знанием сути вопроса и основных свойств объекта.

2. ФОРМУЛЫ ОШИБОК СТАТИСТИЧЕСКИХ ВЕЛИЧИН И СПОСОБЫ ИХ ИСПОЛЬЗОВАНИЯ

Каждая статистическая величина имеет свою статистическую ошибку, с помощью которой можно узнать, насколько выборочная средняя расходится с генеральной средней. Статистические ошибки обозначаются буквой m. Так, например, статистическая ошибка средней арифметической M при большом числе наблюдений выражается так:

$$m_M = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}} \ ,$$

т. е. ошибка прямо пропорциональна изменчивости признака и обратно пропорциональна числу наблюдений. Обычно пользуются приближенной формулой ошибки, а именно:

$$m_M = \frac{\sigma}{\sqrt{n}} \ .$$

Если число наблюдений соответствует малой выборке, то формула ошибки средней арифметической M выглядит так:

$$m_M = \frac{\sigma}{\sqrt{n-1}} \ .$$

Определим ошибку средней арифметической для показателя среднего удоя $M = 5000$ *кг* при $\sigma = 500$ *кг*, при $n = 100$:

$$m_M = \frac{500}{\sqrt{100}} = \frac{500}{10} = 50 \text{ кг}.$$

Принято записывать статистическую характеристику вместе с ее ошибкой. Для данного примера со средней арифметической M это следует записать следующим образом:

$$M \pm m = 5000 \pm 50 \text{ кг}.$$

Смысл ошибки заключается в том, что значение генеральной средней будет находиться в границах утроенной ошибки в ту и другую сторону от выборочной средней, что можно представить следующим образом: генеральная средняя = выборочной средней $\pm 3m$, или для средней арифметической это будет выглядеть так:

$$\overline{M} = M \pm 3m.$$

Это означает для приведенного примера, что генеральная средняя удоя будет находиться в границах $M + 3m$ и $M - 3m$, т. е. $5000 + 3 \cdot 50 = 5150$ *кг* и $5000 - 3 \cdot 50 = 4850$ *кг*.

Значит, чем больше в выборке наблюдений, тем меньше расходятся границы, в которых заключена средняя и, следовательно, тем ближе будет выборочная средняя к генеральной средней.

Так, если бы в приведенном примере n было равно не 100, а 1000 наблюдениям, то величина ошибки уменьшилась бы значительно и была бы равна не 50 *кг*, а 15,81 *кг*

$$\left(m = \frac{500}{\sqrt{1000}} = 15{,}81\ кг.\right).$$

Показателем точности опыта или показателем достоверности (вероятности) того, что выборочная средняя правильно отражает генеральную среднюю, служит так называемый критерий достоверности t, который указывает, сколько раз ошибка помещается в своей средней, т. е. $t = \frac{\text{средняя}}{\text{ошибка средней}}$.

Если t больше или равно 3 ($t \geqslant 3$), то это означает, что выборочная средняя правильно отражает генеральную среднюю, и вероятность этого события $P = 0{,}997$, т. е. если бы сделать 1000 выборок аналогичных той, в которой получены соответствующие показатели средних и их ошибок, то только в трех из них средняя выборочная неправильно бы отражала генеральную среднюю.

Величины критерия достоверности t и вероятности P связаны друг с другом следующим образом:

при $t = 2$	$P = 0{,}955$;
при $t = 2{,}5$	$P = 0{,}988$;
при $t = 3$	$P = 0{,}997$;
при $t = 3{,}5$	$P = 0{,}9995$;
при $t = 4$	$P = 0{,}99994$.

Принято считать, что достаточная точность опыта будет при $t=3$ и $P=0{,}997$. В полевых опытах требование к величине критерия достоверности может быть снижено до $t=2{,}5$ ($P=0{,}988$); при работе, требующей более высокой вероятности, значение t берется не менее 4 ($P = 0{,}99994$). Повышать критерий достоверности особенно важно, если ведется работа по изучению влияния дозы лекарства на организм человека или животного. Положим, в опытах установлено, что доза лекарства в 10 *мг* не вредна и ею можно пользоваться для лечебных целей. Если в опытах для этой дозы получена ошибка $m = \pm 3$ *мг*, то тогда критерий достоверности будет равен $t = \frac{10}{3} = 3{,}3$; это соответствует вероятности $P = 0{,}997$, т. е. из 1000 случаев использования лекарства в такой дозе в лечебных целях в 3 случаях имеет место отравление. Следовательно, в таком случае показатель t, равный 3,3, будет для этой работы недостаточен. Поэтому необходимо увеличить число опытов (наблюдений), подтверждающих безвредность дозы, и при этом получить уменьшенную ошибку для изучаемого показателя дозы и тем самым убедиться в том, что найденная доза может служить массовым лечебным препаратом.

Повышенный *критерий достоверности* потребовалось бы иметь и при установлении уровня безвредных доз облучения ионизирующей радиацией.

Статистические ошибки и критерий достоверности в значительной степени апробируют итоги опытов. Так, например, если в опыте получена ошибка, дающая $t = 2$ (что соответствует $P = 0{,}955$), то это означает, что пользоваться полученными показателями выборочных средних нельзя, так как они в слабой мере отражают величину средней, присущей генеральной совокупности. Поэтому полученные в опыте данные не могут быть распространены на генеральную совокупность по изучаемому явлению, а отсюда и весь опыт носит частный характер, из которого нельзя делать выводов для перенесения их на генеральную совокупность.

Таким образом, вычисление любой средней статистической характеристики или какого-либо коэффициента должно сопровождаться вычислением их статистических ошибок и критерия достоверности. Без этих данных полученные средние не имеют общего научного значения.

Перечислим ошибки различных статистических средних и различных коэффициентов:

а. Ошибка средней арифметической. Эта ошибка чаще всего выражается такой формулой:

$m_M = \frac{\sigma}{\sqrt{n}}$, а при малой выборке — $m_M = \frac{\sigma}{\sqrt{n-1}}$, но можно пользоваться и такими формулами:

$$m_M = \sqrt{\frac{\Sigma pa^2}{n^2}}\,;\quad m_M = \sqrt{\frac{\Sigma V^2}{n} - M^2} \cdot \sqrt{\frac{1}{n-1}}\,.$$

Критерий достоверности для средней арифметической вычисляется по такой формуле:

$$t_M = \frac{M}{m_M} \geqslant 3.$$

б. Определение достоверности средней арифметической при малом числе наблюдений по методу Стьюдента — Фишера. По методу Стьюдента — Фишера критерий достоверности для средней арифметической при малом числе наблюдений определяется табличным методом с учетом так называемого числа степеней свободы n' или ν (ню), зависящего от числа наблюдений в выборке.

Число степеней свободы вычисляется путем уменьшения числа наблюдений на 1, т. е. $\nu = n' = n - 1$. Но в ряде случаев определение n' осуществляется и другим способом, что будет изложено подробно в главе VIII. В таблице Стьюдента — Фишера (см. стр. 124) в верхней строчке указаны величины t при вероятности P; в крайней левой колонке даны значения степеней свободы ν. В зависимости от величины P и ν на пересечении их столбцов и строк в клетках таблицы приведены значения критерия достоверности t. На стр. 124 (см. табл. 16) приводится только часть таблицы, так как из нее исключаются графы, относящиеся к вероятностям, равным 0,80; 0,89, а остаются лишь графы для вероятностей, равных 0,95; 0,98; 0,99 и 0,999[7].

Пример. В четырех повторностях опыта получена средняя озернённость колосьев ржи, равная 25 зёрнам ($M=25$, $n=4$), а сигма равна 5 ($\sigma = 5$).

Определить достоверность средней арифметической.

Число степеней свободы $n' = n - 1 = 4 - 1 = 3$. При таком n' и при высокой вероятности ($P = 0{,}99$) критерий достоверности, определяемый по табл. 16, должен быть равен 5,484. Наши же данные дают следующие значения:

$$m_M = \frac{\sigma}{\sqrt{n-1}} = \frac{5}{\sqrt{3}} = \frac{5}{1{,}732} = 2{,}88,$$

откуда $t = \dfrac{M}{m_M} = \dfrac{25}{2{,}88} \approx 9$. Следовательно, вычисленное значение t больше табличного, поэтому полученная нами средняя арифметическая достоверна. Зная эти данные, можно определить, насколько выборочная средняя отличается от генеральной средней M при заданной величине P ($P = 0{,}99$).

Так как $t = \dfrac{\Delta}{m}$, то отсюда $\Delta = t \cdot m$; в указанном примере $\Delta = 5{,}484 \cdot 2{,}88 = 15{,}8$ зёрнам.

Откуда: $M - \Delta = 25 - 15{,}8 = 9{,}2$ зёрнам; $M + \Delta = 25 + 15{,}8 = 40{,}8$ зёрнам; следовательно, в этих пределах будет находиться значение генеральной средней.

в. Ошибки среднего квадратического отклонения σ **и коэффициента изменчивости** C **выражаются** следующими формулами:

$$m_\sigma = \frac{\sigma}{\sqrt{2n}}; \qquad m_C = \frac{C}{\sqrt{2n}}.$$

---

[7] Полную таблицу можно найти в книге: А. А. Сапегин. Вариационная статистика. М., Сельхозгиз, 1937, стр. 172.

Таблица 16

Таблица Стьюдента—Фишера для определения критерия достоверности

Число степеней свободы, ν	Показатели t при вероятности P			
	$P=0,95$	$P=0,98$	$P=0,99$	$P=0,999$
1	12,71	31,821	63,630	636,20
2	4,30	6,965	9,925	31,60
3	3,18	4,541	5,484	12,94
4	2,78	3,747	4,604	8,61
5	2,57	3,365	4,032	6,86
6	2,45	3,143	3,707	5,96
7	2,36	2,998	3,499	5,40
8	2,31	2,896	3,335	5,04
9	2,26	2,821	3,256	4,78
10	2,23	2,764	3,169	4,59
11	2,20	2,718	3,106	4,49
12	2,18	2,681	3,055	4,32
13	2,16	2,650	3,012	4,12
14	2,14	2,624	2,977	4,14
15	2,13	2,602	2,947	4,07
16	2,12	2,583	2,921	4,02
17	2,11	2,567	2,898	3,96
18	2,10	2,552	2,878	3,92
19	2,09	2,539	2,865	3,88
20	2,09	2,528	2,845	3,85
21	2,08	2,518	2,831	3,82
22	2,07	2,508	2,819	3,79
23	2,07	2,500	2,807	3,77
24	2,06	2,492	2,797	3,75
25	2,06	2,485	2,787	3,72
26	2,06	2,479	2,779	3,71
27	2,05	2,473	2,771	3,69
28	2,05	2,467	2,763	3,67
29	2,05	2,462	2,750	3,66
30	2,04	2,457	2,750	3,64

Критерии достоверности этих величин соответственно:

$$t = \frac{\sigma}{m_\sigma} \geqslant 3 \text{ и } t_C = \frac{C}{m_C} \geqslant 3.$$

г. Ошибка разности между средними арифметическими двух выборок (для коррелированных и некоррелированных выборок при большом и малом числах наблюдений). В экспериментальной работе большое значение приобретает определение ошибки разности между средними арифметическими, так как конечная цель опыта чаще всего сводится к выяснению, имеет ли место разница в показателях средних арифметических между объектами контрольной и подопытной групп. При этом производится сравнение величин средних арифметических опыта и контроля. Получаемая разность $D = M_{оп} - M_{контр}$ и показывает исследователю, имел ли место сдвиг в изучаемых показателях или нет.

Для выяснения достоверности этой разности требуется вычислить величину ее ошибки m_D с помощью следующей формулы:

$$m_D = \sqrt{m_1^2 + m_2^2},$$

где m_1 — ошибка средней арифметической первой выборки (опыт);

m_2 — ошибка средней арифметической второй выборки (контроль).

Этой формулой пользуются в тех случаях, *когда выборки между собой некоррелированы*, т. е. варианты одной выборки не связаны и не зависят от величины вариантов другой. Примером таких некоррелированных выборок, по которым идет сравнение их средних арифметических, может служить средняя урожайность посевов ржи двух сортов или средние показатели плодовитости облученных и необлученных мышей. Проведем вычисление ошибки разности для последнего примера. *Предположим, что показатели по группам мышей были следующими* (табл. 17).

Определить, достоверно ли снижение плодовитости самок мышей в среднем на 2 головы под влиянием облучения.

Из табл. 17 видно, что ошибка средней в группе облученных мышей равна 0,14 голов, а в контрольной 0,08 голов. Подставляем квадраты этих ошибок в формулу ошибки разности m_D:

$$m_D = \sqrt{m_1^2 + m_2^2} = \sqrt{0{,}14^2 + 0{,}08^2} + \sqrt{0{,}0196 + 0{,}0064} =$$
$$= \sqrt{0{,}0260} = 0{,}16 \text{ голов.}$$

Отсюда $D \pm m_D = 2 \pm 0,16$ голов. Критерий достоверности t_D, вычисляемый обычным способом, оказывается равным $\frac{D}{m_D} = \frac{2}{0,16} \approx 12$. Так как достаточно иметь критерий достоверности не меньше 3, чтобы считать статистическую величину достоверной, то в данном примере полученное значение $t \approx 12$ указывает на полную достоверность того, что облучение вызвало снижение плодовитости в среднем на 2 мышонка в помете.

Таблица 17

Определение достоверности разности между средними арифметическими двух некоррелированных выборок

Группа мышей	Число самок, n	Средняя плодовитость (числа голов), M	σ	$m_M = \frac{\sigma}{\sqrt{n}}$
Облученные . . .	49	6	1,0	0,14
Контрольные . . .	45	8	0,5	0,08
	—	$D = 2$ (разность)	—	$m_D = 0,16$

В исследованиях довольно часто встречаются такие случаи, когда выборки коррелируют между собой, т. е. когда значения вариантов одной выборки зависят от значений вариантов другой выборки.

Примером коррелированных выборок могут служить также данные, характеризующие родительское и дочернее потомство в генетических работах. Так, например, известно, что удои коров-дочерей в какой-то мере зависят (коррелированы) от уровня удоя их матерей, что и должно учитываться при обработке этих данных.

Для коррелированных выборок формула ошибки разности m_D между средними арифметическими двух выборок несколько усложняется по сравнению с формулой m_D для некоррелированных выборок и выглядит следующим образом:

$$m_D = \sqrt{m_1^2 + m_2^2 - 2m_1 \cdot m_2 \cdot r}.$$

Здесь m_1 и m_2 — ошибки средних арифметических по обеим выборкам; r — коэффициент корреляции между вариантами обеих выборок.

Введение в формулу выражения $2m_1 \cdot m_2 \cdot r$ уменьшает показатель ошибки разности коррелированных выборок, что дает более высокий критерий достоверности t. Но следует иметь в виду, что если не представляется возможным вычислить значение r, то и для коррелированных выборок в таком случае можно пользоваться формулой $m_D = \sqrt{m_1^2 + m_2^2}$, но при этом показатель ошибки разности будет несколько завышен, что снизит уровень достоверности разности.

Разберем пример с вычислением достоверности разности для коррелированных выборок.

Пример. Определить, достоверно ли селекционное заключение о том, что бык улучшает продуктивность своих коров-дочерей, если удои дочерей в среднем на 500 *кг* выше, чем удои их матерей, а коэффициент корреляции, показывающий зависимость удоя дочерей от удоя их матерей, равен $r = +0{,}4$. Число обследованных дочерей у быка равно 100, среднее квадратическое отклонение по удоям дочерей равно 500 *кг*, а по удоям их матерей — 400 *кг*.

Сводные показатели по этому примеру приведены в табл. 18.

Таблица 18

Определение достоверности разности между средними арифметическими двух коррелированных выборок

Группа коров	Число сравниваемых пар, n	Средний удой в *кг*, M	σ	$m_M = \frac{\sigma}{\sqrt{n}}$
Дочери	100	4500	500	50
Матери	100	4000	400	40
		$D = 500$ (разность)	—	$m_D = 50$

Подставим в формулу ошибки разности необходимые значения:

$$m_D = \sqrt{m_1^2 + m_2^2 - 2 \cdot m_1 \cdot m_2 \cdot r} = \sqrt{50^2 + 40^2 - 2 \cdot 50 \cdot 40 \cdot 0{,}4} =$$

$$= \sqrt{2500 + 1600 - 1600} = \sqrt{2500} = 50 \text{ кг},$$

откуда:

$$D \pm m_D = 500 \pm 50; \quad t_D = \frac{500}{50} = 10.$$

Значение критерия достоверности $t = 10$ указывает на полную достоверность того, что испытуемый производитель в среднем повышает продуктивность своего потомства на 500 *кг* молока по сравнению с уровнем продуктивности матерей потомства.

При малой выборке ($n < 30$), что особенно часто встречается в физиологических и острых генетических опытах, определение достоверности разности осуществляется при более высоком уровне t. Метод определения достоверности разности при малом числе наблюдений разработан Стьюдентом и Фишером и позволяет значительно расширить использование статистической обработки в экспериментальных исследованиях. Для этих целей пользуются специальными таблицами, в которых приводится достоверное значение t при различном числе наблюдений в выборке и при различных уровнях вероятности P. Удобная рабочая таблица, приведенная в книге Н. А. Плохинского, помещена в табл. 19.

Как видно из ее построения, достоверность разницы можно определять для коррелированных и некоррелированных выборок. Для первого случая в правой крайней колонке приводятся данные о числе сравниваемых пар n, вошедших в коррелированную выборку, а в крайней левой колонке — сведения о суммарном числе наблюдений $(n_1 + n_2)$ в сравниваемых некоррелированных выборках. Остальные три колонки заполняются данными о значении критерия достоверности t. Вторая колонка слева показывает достоверное значение t при высоком уровне вероятности $P = 0{,}997$, третья колонка слева дает значение t при среднем уровне вероятности $P = 0{,}99$ и в следующей колонке значение t соответствует наиболее низкому уровню вероятности $P = 0{,}95$, используемому при полевых и производственных опытах.

Разберем на примере использование таблицы.

Пример. Физиологические опыты по изучению влияния температуры на частоту дыхательных движений у коров показали, что при температуре воздуха в 25°С среднее число дыхательных движений равно 45, а при температуре воздуха в 10°С оно снижается до 30. Опыт проведен на одной и той же группе коров численностью в 5 голов, т. е. выборки коррелированы и коэффициент $r = 0{,}80$.

Статистические данные после обработки были получены следующие (табл. 20).

Из вычисленных значений t для средних арифметических в каждой группе опыта видно, что средние показатели ча-

Таблица 19

Критерий достоверности при малых выборках по Стьюденту—Фишеру

Выборки некоррелированы. Число наблюдений в обеих выборках (n_1+n_2)	Критерий достоверности разности по требованиям точности			Выборки коррелированы. Число сравниваемых пар, n
	высокому $P=0{,}997$	среднему $P=0{,}99$	низкому $P=0{,}95$	
	3,0	2,5	2,0	
$(n_1+n_2)>32$				$n>31$
32	3,05	2,75	2,04	31
31	3,10	2,76	2,05	30
30	3,15	2,76	2,05	29
29	3,20	2,77	2,05	28
28	3,25	2,78	2,06	27
27	3,30	2,79	2,06	26
26	3,35	2,80	2,06	25
25	3,40	2,81	2,07	24
24	3,45	2,82	2,07	23
23	3,50	2,83	2,08	22
22	3,55	2,85	2,09	21
21	3,57	2,86	2,09	20
20	3,60	2,88	2,10	19
19	3,65	2,90	2,11	18
18	3,70	2,92	2,12	17
17	3,75	2,95	2,13	16
16	3,80	2,98	2,15	15
15	3,85	3,01	2,16	14
14	3,95	3,06	2,18	13
13	4,00	3,11	2,20	12
12	4,15	3,17	2,23	11
11	4,30	3,25	2,26	10
10	4,50	3,36	2,31	9
9	4,80	3,50	2,37	8
8	5,20	3,71	2,45	7
7	5,90	4,03	2,57	6
6	7,58	4,60	2,78	5
5	10,24	5,84	3,18	4
4	22,38	9,93	4,30	3
3	318,54	63,66	12,71	2

стоты дыхания имеют очень большую вероятность, так как в первой группе $t = 15$, а во второй $t = 7{,}5$, т. е. значительно больше 3, когда вероятность $P = 0{,}997$. Для определения достоверности разницы воспользуемся табл. 19 на стр. 129.

Таблица 20

Определение достоверности разницы между средними арифметическими двух коррелированных групп при малом числе наблюдений

Температурные различия в °C	Число животных в группе	Среднее число дыхательных движений, M	Статистическая ошибка, m_D	Критерий достоверности, t
25 10	5 5	45 30	3 4	15 7,5
	—	$D = -15$ (разность)	$m_D = \sqrt{m_1^2 + m_2^2 - 2m_1 \cdot m_2 \cdot r} = \sqrt{3^2 + 4^2 - 2 \cdot 3 \cdot 4 \cdot 0{,}8} = \sqrt{5{,}80} = 2{,}41$	6,2

Численность n в сравниваемых коррелированных выборках в рассматриваемом примере равна 5 парам. В правой крайней колонке находим значение $n = 5$ и по строчке определяем достоверное значение t. Для разных уровней вероятности оно оказывается следующим: $t = 7{,}58$ при $P = 0{,}997$; $t = 4{,}60$ при $P = 0{,}99$ и $t = 2{,}78$ при $P = 0{,}95$. Вычисленное нами значение t для разности средних арифметических равно 6,2. Следовательно, снижение частоты дыхания при снижении температуры окружающего воздуха в среднем на 15 дыхательных движений достоверно при низких и средних уровнях вероятности $P = 0{,}95$, $P = 0{,}99$, что вполне достаточно для требований достоверности к физиологическим опытам такого типа.

Аналогичным образом определяется достоверность разности между средними арифметическими двух некоррелированных выборок.

Пример. Определить, достоверна ли разница в потреблении кислорода коровами джерсейской и остфризской пород, если в результате опыта и его статистической обработки получены следующие данные (табл. 21).

Суммарная численность n обеих выборок равна $n_1 + n_2 = 10 + 8 = 18$ При таком значении n, взятом в крайней левой колонке для некоррелированных выборок, достоверная величина $t = 3,70$ (при $P = 0,997$), $t = 2,92$ (при $P = 0,99$) и $t = 2,12$ (при $P = 0,95$) Вычисленное же в приведенном примере значение $t = 2,5$ Следовательно, при низком уровне вероятности, достаточном при проведении физиологических опытов в производственных условиях, полученная породная разница в интенсивности газообмена у коров двух пород статистически достоверна

Таблица 21

Определение достоверности разницы между средними арифметическими двух некоррелированных групп при малом числе наблюдений

Порода животных	Число животных	Поглощено кислорода *л* на *кг* веса в *час*, M	Ошибка, m	t
Джерсейская	10	0,541	0,050	10,8
Остфризская	8	0,335	0,065	5,1
	—	$D = 0,206$ (разность)	0,082	2,5

д. Ошибка доли при альтернативных признаках. Если выборка составлена из объектов, у которых варьирующий признак имеет качественные отличия (и даже альтернативные), то в этих случаях вместо средней арифметической средними показателями служат доля особей, имеющих данный качественный признак, и доля особей, у которых этот признак отсутствует Типичным примером альтернативных признаков является доля особей мужского пола и женского пола в потомстве каких-либо животных

Предположим, что в потомстве одной самки наблюдается существенный сдвиг пола в сторону более частого рождения самцов При этом из общего числа детенышей, рожденных самкой, равного 150 головам, 80% особей мужского пола и только 20% — женского [или в долях единицы это может быть выражено как 0,8 (x) и 0,2 (y)]

Ошибка доли рождения приплода каждого пола выражается формулой

$$m_x = m_y = \sqrt{\frac{x\ y}{n}} = \sqrt{\frac{x(1-x)}{n}}\ .$$

То есть, ошибка доли каждого из признаков равна корню квадратному из произведения долей по каждому признаку, деленному на число наблюдений.

В рассматриваемом примере при $n = 150$ рождено самцов 120 и самок 30. Доля рожденных самцов составляет $x = \frac{120}{150} = 0{,}8$, а доля самок $y = \frac{30}{150} = 0{,}2$.

Отсюда ошибка доли рождения самцов равна:

$$m_x = m_y = \sqrt{\frac{x \cdot y}{n}} = \sqrt{\frac{0{,}8 \cdot 0{,}2}{150}} = \sqrt{0{,}001} = 0{,}033.$$

Таким образом, достоверность доли рождения самцов $t_x = \frac{x}{m_x} = \frac{0{,}8}{0{,}033} = 24$ и достоверность для доли рождения самок $t_y = \frac{y}{m_y} = \frac{0{,}2}{0{,}033} = 6$, что указывает на полную достоверность доли рождения как самцов, так и самок.

Зная ошибку доли в выборке, можно подсчитать, какое соотношение полов будет иметь место и в генеральной совокупности.

Так, доля самцов, рождающихся в генеральной совокупности, может находиться в пределах $\pm 3\,m_x$ от выборочной доли рождения самцов.

Следовательно, из данных примера это будет равно

$$\overline{x} = x \pm 3m_x \quad \text{или} \quad \overline{x} = 0{,}8 \pm 3 \cdot 0{,}033,$$

где $\overline{x}$ — доля генеральной совокупности.

Отсюда:

$$\overline{x} = 0{,}8 + 0{,}099 = 0{,}899;$$

$$\overline{x} = 0{,}8 - 0{,}099 = 0{,}701,$$

т. е. в генеральной совокупности рождение самцов колеблется в пределах от 70,1% до 89,9% всего рождающегося потомства.

е. Ошибка разности долей. Для вычисления достоверности разности долей, определяющей встречаемость особей с одинаковым альтернативным признаком в двух выборках, пользуются такой же формулой ошибки, как и в случае определения ошибки разности между средними арифметическими при количественных признаках. В этом случае формула ошибки разности долей следующая:

$$m_d = \sqrt{m_x^2 + m_y^2},$$

где $d = x - y$ — разность долей;
m_x — ошибка доли по признаку x;
m_y — ошибка доли по отсутствию признака x.

Пример. В контрольной группе самок доля рождения самцов составляла 0,5; в потомстве же облученных самок — 0,7.

Определить, достоверно ли увеличенное рождение самцов в группе облученных самок, если разница долей с контрольной группой составляет $d = 0{,}7 - 0{,}5 = 0{,}2$, а число подопытных самок в контрольной группе равно 49 и в облученной — 50.

Находим ошибку доли по каждой группе самок.

В облученной группе:

$$m_x \text{ (облуч.)} = \sqrt{\frac{0{,}7 \cdot 0{,}3}{50}} = \sqrt{\frac{0{,}21}{50}} = \sqrt{0{,}0042} = 0{,}065.$$

В контрольной группе:

$$m_x \text{ (контр.)} = \sqrt{\frac{0{,}5 \cdot 0{,}5}{49}} = \sqrt{\frac{0{,}25}{49}} = \sqrt{0{,}0051} = 0{,}07.$$

Разность долей

$$d = x_{\text{облуч}} - x_{\text{контр}} = 0{,}7 - 0{,}5 = 0{,}2.$$

Ошибка разности

$$m_d = \sqrt{m_x^2 \text{ (облуч.)} + m_x^2 \text{ (контр.)}} = \sqrt{0{,}065^2 + 0{,}07^2} = 0{,}095,$$

откуда:

$$t_d = \frac{d}{m_d} = \frac{0{,}2}{0{,}095} = 2{,}1.$$

Такое значение **t** при пониженном требовании к вероятности указывает на достоверность сдвига в соотношении полов под влиянием облучения.

ж. Ошибка разности между выборочной и генеральной долями. В генетических работах довольно часто приходится сравнивать данные, полученные экспериментальным путем, с данными, которые характеризуют теоретическую изменчивость признака, типичную для генеральной совокупности.

Так, например, положим, что на 500 гибридов второго поколения при моногибридном скрещивании получено 60% особей с доминантным и 40% — с рецессивным признаками. В то же время известно, что для альтернативных менделирующих признаков типично такое расщепление во втором поколении гибридов: 75% особей с доминантным и 25% — с рецессивным признаками, что и будет составлять расщепление, характерное для генеральной совокупности. Таким образом, в данном примере разность между генеральной и выборочной долями будет составлять $d = 0{,}75 - 0{,}6 = 0{,}15$.

Определим достоверность разности генеральной и выборочной долей.

Для этого используем известную уже формулу ошибки разности

$$m_d = \sqrt{m^2_{\substack{\text{выбор}\\ \text{доли}}} + m^2_{\substack{\text{генер}\\ \text{доли}}}}\,.$$

Но так как генеральная доля не может иметь ошибки репрезентативности, т. е. статистической ошибки, создающейся за счет сравнения части с целым (т. е. выборочных показателей с генеральной совокупностью), ибо генеральная доля нацело отражает эту закономерность в генеральной совокупности, то формула ошибки разности будет упрощаться, принимая следующий вид:

$$m_D = \sqrt{m^2_{\substack{\text{выбор}\\ \text{доли}}} + 0} = m_{\substack{\text{выбор.}\\ \text{доли}}}$$

Иначе говоря, ошибка разности выборочной и генеральной долей равна ошибке выборочной доли.

Критерий достоверности определяется обычным путем и выражается следующим образом:

$$t_d = \frac{d}{m_{\substack{\text{выбор.}\\ \text{доли}}}} = \frac{d}{m_d}\,.$$

В нашем примере $d = 0{,}15$, а ошибка $m_D = m_{\text{выбор. доли}}$. Вычислим ошибку выборочной доли по приведенной на стр. 131 формуле:

$m_x = \sqrt{\frac{x \cdot y}{n}}$, где x — доля особей с доминантным признаком; y — доля особей с рецессивным признаком; n — число особей в выборке. Отсюда ошибка доли для особей с доминантным признаком в указанном примере равна:

$$m_{\iota} = \sqrt{\frac{0,6 \cdot 0,4}{500}} = \sqrt{\frac{0,24}{500}} = \sqrt{0,00048} = 0,022,$$

и этот же результат будет служить показателем ошибки разности d. Отсюда:

$$d \pm m_d = 0,15 \pm 0,022; \quad t = \frac{0,15}{0,022} \approx 7.$$

Значение $t = 7$ указывает на достоверность разницы между долями доминантных особей генеральной и выборочной совокупностей. Следовательно, уменьшение доли особей с доминантным признаком не случайно, а обусловлено воздействием каких-либо факторов, сдвигающих соотношение в расщеплении гибридов второго поколения.

3. Ошибка коэффициента корреляции для больших и малых выборок (метод z). Ошибка коэффициента корреляции при большом числе наблюдений выражается формулой:

$$m_r = \frac{1 - r^2}{\sqrt{n}},$$

где r — выборочный коэффициент корреляции; n — число наблюдений. В этом случае большим числом наблюдений считается $n \geqslant 100$.

Критерий достоверности определяется уже известным способом:

$$t_r = \frac{r}{m_r} \geqslant 3.$$

Значение коэффициента корреляции генеральной совокупности находится в пределах $\pm$ утроенной ошибки, а именно:

$$r_{\text{генер}} = r_{\text{выбор}} \pm 3m_r.$$

Если число наблюдений в выборке меньше 100, то критерий достоверности t может быть использован только с учетом *численности выборки и должен иметь значение* $t > 3$, чтобы подтвердить достоверность коэффициента корреляции, вычисленную для малой выборки.

Для определения достоверности r при малых выборках пользуются так называемым «методом зет» (Z), предложенным Фишером, или формулой $m = \sqrt{\frac{1 - r^2}{n - 2}}$.

Величина Z является функцией коэффициента корреляции, что выражается следующим уравнением:

$$Z = r + \frac{1}{3}r^3 + \frac{1}{5}r^5 + \frac{1}{7}r^7 + \ldots$$

Так как ошибка величины выборочного Z не зависит от величины Z генеральной совокупности, то ошибка выборочного Z выражается формулой, простой для вычисления, а именно:

$$m_z = \frac{1}{\sqrt{n-3}},$$

а критерий достоверности Z

$$t_z = \frac{z}{m_z} = Z\cdot\sqrt{n-3}$$

достоверен при определенной численности выборки.

Для каждого значения Z можно определить его достоверность при определенном числе наблюдений n, если воспользоваться приведенной выше формулой: $t_z = 3 = Z\sqrt{n-3}$. Отсюда $n = \frac{9}{z^2} + 3$. На основании этого вычисляется значение n, дающее достоверную величину Z, которой будет соответствовать и определенное значение коэффициента корреляции r. Подсобные таблицы, указывающие достоверные значения Z и r при определенном числе наблюдений n, позволяют быстро определить достоверность вычисленного в конкретном случае значения коэффициента корреляции (см. табл. 22 на стр. 137).

Пользуются этой таблицей следующим образом.

Например, в опыте на 20 коровах изучалась связь между частотой дыхания и уровнем температуры окружающего воздуха. Оказалось, что эта связь выразилась коэффициентом корреляции, равным $r = 0{,}65$. Определить достоверность величины r.

Для этого по табл. 22 находим, что значение r будет достоверно при 18 наблюдениях, а в примере $n = 20$, следовательно, полученное значение $r = 0{,}65$ при такой численности выборки вполне достоверно, так как имеющееся число наблюдений больше того числа наблюдений, которое имеется в таблице при данной величине r, т. е. $n_{выбор.} > n_{табл.}$ $(20 > 18)$.

Таблица 22

Предельная численность выборки, достаточная для достоверности коэффициента корреляции и величины Z (при t = 3) для больших и малых выборок (из книги Н. А. Плохинского)

r	Z	n	r	Z	n	r	Z	n	r	Z	n	r	Z	n
0,01	0,010	90003	0,21	0,213	201	0,41	0,435	51	0,61	0,71	21	0,81	1,12	10
0,02	0,020	22503	0,22	0,224	183	0,42	0,448	48	0,62	0,72	20	0,82	1,15	10
0,03	0,030	10003	0,23	0,234	167	0,43	0,460	46	0,63	0,74	19	0,83	1,18	9
0,04	0,040	5628	0,24	0,245	153	0,44	0,472	44	0,64	0,75	19	0,84	1,22	9
0,05	0,050	3603	0,25	0,255	141	0,45	0,485	41	0,65	0,77	18	0,85	1,25	9
0,06	0,060	2503	0,26	0,266	130	0,46	0,497	39	0,66	0,79	17	0,86	1,29	8
0,07	0,070	1840	0,27	0,277	120	0,47	0,510	38	0,67	0,81	17	0,87	1,33	8
0,08	0,080	1409	0,28	0,288	112	0,48	0,523	36	0,68	0,82	16	0,88	1,37	8
0,09	0,090	1114	0,29	0,299	104	0,49	0,536	34	0,69	0,85	15	0,89	1,42	7
0,10	0,103	894	0,30	0,309	97	0,50	0,549	33	0,70	0,87	15	0,90	1,47	7
0,11	0,110	741	0,31	0,321	91	0,51	0,563	31	0,71	0,89	14	0,91	1,52	7
0,12	0,121	624	0,32	0,332	85	0,52	0,576	30	0,72	0,91	14	0,92	1,58	7
0,13	0,131	529	0,33	0,343	80	0,53	0,598	29	0,73	0,93	13	0,93	1,65	6
0,14	0,141	456	0,34	0,354	75	0,54	0,604	28	0,74	0,95	13	0,94	1,72	6
0,15	0,151	400	0,35	0,365	71	0,55	0,618	27	0,75	0,97	12	0,95	1,81	6
0,16	0,161	348	0,36	0,377	67	0,56	0,633	25	0,76	0,99	12	0,96	1,95	5
0,17	0,172	308	0,37	0,388	63	0,57	0,647	24	0,77	1,02	12	0,97	2,09	5
0,18	0,182	281	0,38	0,400	59	0,58	0,663	23	0,78	1,04	11	0,98	2,29	5
0,19	0,192	247	0,39	0,412	56	0,59	0,677	23	0,79	1,07	11	0,99	2,64	4
0,20	0,203	222	0,40	0,427	53	0,60	0,693	22	0,80	1,09	10	0,999	3,80	4

и. Ошибки репрезентативности для других показателей связи (корреляционного отношения η, полихорического показателя связи ρ, коэффициента регрессии R, коэффициента корреляции при альтернативных признаках r_A) выражаются следующими формулами

ошибка корреляционного отношения

$$m_\eta = \frac{1-\eta^2}{\sqrt{n}} \quad \text{при} \quad t \geqslant 4,$$

ошибка полихорического показателя связи

$$m_\rho = 2\sqrt{\frac{\alpha(\alpha-1)}{n(l_1-1)(l_2-1)}};$$

ошибки коэффициентов регрессии

$$m_{R_{xy}} = \frac{\sigma_x}{\sigma_y}\sqrt{\frac{1-r^2}{n}} \quad \text{и} \quad m_{R_{yx}} = \frac{\sigma_y}{\sigma_x}\sqrt{\frac{1-r^2}{n}} \quad \text{при} \quad t \geqslant 4$$

(здесь σ_x и σ_y — средние квадратические отклонения по каждому признаку, r — коэффициент корреляции между признаками x и y, n — число наблюдений),

ошибка коэффициента корреляции для альтернативных признаков

$$m_{r_A} = \frac{1-r_A^2}{\sqrt{n}};$$

ошибка критерия криволинейности

$$m_L = 2\sqrt{\frac{L^2}{n}} \cdot \sqrt{(1-\eta^2)^2-(1-r^2)^2+1}.$$

3. МЕТОД ХИ-КВАДРАТ (χ^2) КАК МЕРА ВЕРОЯТНОСТИ СОВПАДЕНИЯ ЧАСТОТ РАЗНЫХ ВАРИАЦИОННЫХ РЯДОВ

(эмпирического и теоретического)

При обработке материалов возникает необходимость сопоставления частот одного ряда с частотами другого ряда и выяснения вероятности того, что частоты обоих рядов несущественно расходятся между собой Такая постановка вопроса особенно часто встает при необходимости сравнения частот эмпирического ряда с частотами теоретического ряда, которые отражают закономерность генеральной совокупности

Для этих целей используются так называемые критерии согласия хи-квадрат (χ^2), предложенный Е С Пирсоном и лямбда (λ), предложенный А Н Колмогоровым

Разберем применение в указанных целях критерия согласия хи-квадрат Пирсона

Формула хи-квадрат выражается суммой частных величин, полученных путем деления квадрата отклонения эмпирических частот от теоретических частот на теоретические частоты

$$\chi^2 = \Sigma \frac{(p_{\text{эмпир}} - p_{\text{теорет}})^2}{p_{\text{теорет}}} .$$

Значение хи-квадрат изменяется от 0 до ∞ Если оно равно 0, то эмпирические частоты совпадают с теоретическими частотами, а чем больше они отклоняются от теоретических частот, тем больше значение хи-квадрат Каждому значению хи-квадрат соответствует определенная вероятность P, при этом, чем больше хи квадрат, тем меньше вероятность его получения

Для суждения о совпадении или несовпадении эмпирических частот с теоретическими пользуются специальными таблицами вероятностей P для критерия χ^2 В сокращенном виде такая таблица приведена на стр 140 (см табл 23) [8]

Для пользования таблицей необходимо иметь значение хи-квадрат, вычисленное на основании сопоставления частот двух рядов, а также значение ню (ν), называемое числом степеней свободы Число степеней свободы определяет число классов в вариационном ряду, исправленное на число факторов, вызывающих ограничение варьирования, обусловленных особенностями каждого данного ряда

Чаще всего число степеней свободы равно числу наблюдений или числу классов, уменьшенным на единицу ($\nu = n - 1$) или на какое-то другое число Подробнее это разобрано при рассмотрении конкретного примера в главе VIII

Ознакомимся на примере, разбираемом В И Романовским, с вычислением хи квадрат и определим вероятность расхождения эмпирических и теоретических частот [9]

Пример На 840 делянках был произведен посев пшеницы для определения урожайности В табл 24 (см стр 141) представлен вариационный ряд изменчивости урожайности по делянкам, вычислены теоретические частоты по уравнению нормальной кривой и определено значение хи-квадрат

---

[8] Для подробного рассмотрения метода хи квадрат целесообразно воспользоваться книгой П Ф Рокицкий Основы вариационной статистики для биологов Минск, 1961

[9] В И Романовский Применения математической статистики в опытном деле М—Л, 1947, стр 26

Таблица χ^2 в зависимости от числа степеней свободы (ν) в интервале от 1 до 30 при пяти градациях вероятности P (из книги Ф. Миллса)

Число степеней свободы, ν	Вероятности				
	$P=0,01$	$P=0,05$	$P=0,5$	$P=0,95$	$P=0,99$
1	0,0001	0,003	0,455	3,84	6,63
2	0,0201	0,103	1,386	5,99	9,21
3	0,115	0,352	2,366	7,81	11,34
4	0,297	0,711	3,357	9,49	13,28
5	0,554	1,145	4,351	11,07	15,09
6	0,872	1,635	5,348	12,59	16,81
7	1,239	2,167	6,342	14,07	18,47
8	1,646	2,733	7,344	15,51	20,09
9	2,088	3,325	8,343	16,92	21,67
10	2,558	3,940	9,342	18,31	23,21
11	3,053	4,575	10,341	19,67	24,72
12	3,571	5,226	11,340	21,03	26,22
13	4,107	5,892	12,340	22,37	27,69
14	4,660	6,571	13,339	23,68	29,14
15	5,229	7,261	14,339	25,00	30,58
16	5,812	7,962	15,338	26,30	32,00
17	6,408	8,672	16,338	27,59	33,41
18	7,015	9,390	17,338	28,87	34,80
19	7,633	10,117	18,338	30,14	36,19
20	8,260	10,851	19,337	31,41	37,57
21	8,897	11,591	20,337	32,67	38,93
22	9,542	12,338	21,337	33,92	40,29
23	10,196	13,091	22,337	35,17	41,63
24	10,856	13,848	23,337	36,41	42,98
25	11,524	14,611	24,337	37,65	44,31
26	12,198	15,379	25,336	38,88	45,64
27	12,879	16,151	26,336	40,11	46,96
28	13,565	16,928	27,336	41,34	48,29
29	14,256	17,708	28,336	42,56	49,59
30	14,953	18,493	29,336	43,77	50,89

Теоретические частоты вычисляются уже изложенным на стр. 58, 59 способом по уравнению нормальной кривой:

$$p_x = \frac{n \cdot k}{\sigma \sqrt{2\pi}} \cdot e^{-\frac{(x-\overline{x})^2}{2\sigma^2}} .$$

Вычисление значений частот эмпирического ряда дает в данном примере значения: $\overline{x} = M = 114{,}32$ *г*; $\sigma = 11{,}389$ *г*; $n = 840$; $\kappa = 7$ (классовый промежуток). Эти величины получаются путем обычной обработки вариационного ряда, изложенного в главе IV.

Окончательно уравнение для выражения значений теоретических частот на основании значений n, σ и M эмпирического ряда запишется так:

$$p_x = \frac{840 \cdot 7}{11{,}389\sqrt{2\pi}} \cdot e^{-\frac{(x-114{,}32)^2}{2\ 11{,}389^2}} .$$

Таблица 24

Вычисление величины χ^2 для вариационного ряда с количественным признаком

Середина класса варьирующего признака, x	Эмпирические частоты, $p_{эмпир}$	Теоретические частоты, $p_{теорет}$	Отклонения эмпирических частот от теоретических, $p_{эмпир} - p_{теорет}$	$\chi^2 = \Sigma \frac{(p_{эмпир} - p_{теорет})^2}{p_{теорет}}$
70	1 } 12	0,1 } 7		
77	4	0,9	5,0	$5{,}0^2 : 7{,}0 = 3{,}57$
84	7	6,0		
91	19	25,3	—6,3	$(-6{,}3)^2 : 25{,}3 = 1{,}57$
98	72	73,6	—1,6	$(-1{,}6)^2 : 73{,}6 = 0{,}03$
105	141	147,1	—6,1	$(-6{,}1)^2 : 147{,}1 = 0{,}25$
112	201	201,9	—0,9	$(-0{,}9)^2 : 201{,}9 = 0{,}00$
119	203	189,3	13,7	$13{,}7^2 : 189{,}3 = 0{,}99$
126	125	121,2	3,8	$3{,}8^2 : 121{,}2 = 0{,}12$
133	54	53,7	0,3	$0{,}3^2 : 53{,}7 = 0{,}02$
140	9 } 13	16,4 } 20,4		
147	2	3,4	—7,4	$(-7{,}4)^2 : 20{,}4 = 2{,}68$
154	1	0,5		
161	1	0,1		
Сумма	840	839,5		9,23

По этому равенству, подставляя значения середины классов (x), можно вычислить теоретические частоты $p_{\text{теорет}}$, проставленные в третьей колонке слева табл. 24. Две последние колонки этой таблицы дают для каждого класса соответственно значения ($p_{\text{эмпир}} - p_{\text{теорет}}$) и $[(p_{\text{эмпир}} - p_{\text{теорет}})^2 : p_{\text{теорет}}]$. Суммируя данные последней колонки, получаем величину хи-квадрат, равную 9,23.

Следует иметь в виду, что необходимо соблюдать условия, заключающиеся в том, что в эмпирическом вариационном ряду ни один класс не должен иметь число наблюдений меньше 10[9a]. Если это имеет место, то такие классы объединяют по показателю частот вместе, что и было сделано для первых трех и последних четырех классов вариационного ряда разбираемого примера. Такое объединение уменьшило число классов с 14 до 9.

Определим число степеней свободы для данного примера, зависимого от числа классов, равного в разбираемом примере 9, и от двух ограничений, обусловленных тем, что теоретические частоты вычислялись по двум заданным статистическим величинам $\bar{x}$ и σ, входящим в уравнение нормального распределения; поэтому $\nu = 9 - 2 = 7$.

Если число степеней свободы меньше 30, то для определения вероятности расхождения частот пользуются таблицей Фишера, которая в сокращенном виде приведена на стр. 140 (табл. 23).

В зависимости от того, на каком уровне точности будет определяться значение χ^2, выбирают нужную графу табл. 23.

В крайней левой колонке берется та строчка, которая соответствует числу степеней свободы ню (ν). На пересечении соответствующей строчки и столбца стоит предельное значение хи-квадрат при данном уровне вероятности.

Уровень точности может быть выбран повышенным ($P = 0{,}999$; $P = 0{,}99$; $P = 0{,}95$), средним ($P = 0{,}5$) и пониженным ($P = 0{,}3$; $P = 0{,}2$). Самый низкий допускаемый предел вероятности может быть взят на уровне $P = 0{,}05$.

Если вычисленное значение хи-квадрат, по данным примера, меньше предельного значения хи-квадрат, установленного табл. 23 в соответствии с числом степеней свободы и уровнем допускаемой вероятности, то можно сделать заключение, что данный эмпирический ряд будет соответствовать характеру распределения частот теоретического ряда.

[9a]. Другие авторы рекомендуют ограничение не меньше 5.

Если же вычисленное значение χ^2 больше табличного, то частоты эмпирического ряда не случайно отклоняются от частот теоретического ряда

В приведенном примере вычисленное значение $\chi^2 = 9{,}23$, а число степеней свободы $\nu = 7$ Отсюда по табл 23 при высоких уровнях вероятности ($P = 0{,}95$, $P = 0{,}99$) предельные значения χ^2 равны соответственно 14,07, 18,47

Следовательно, вычисленное значение хи-квадрат для всех высоких уровней вероятности меньше предельного значения хи квадрат, что указывает на соответствие эмпирического ряда теоретическому

Если в нашем распоряжении нет таблицы предельного значения хи квадрат, то критерий достоверности хи квадрат можно вычислить по формуле

$$t_{\chi^2} = \frac{\chi^2 - \nu}{\sqrt{2\nu}} .$$

Воспользуемся этой формулой для разбираемого примера

$$t_{\chi^2} = \frac{\chi^2 - \nu}{\sqrt{2\nu}} = \frac{9\,23 - 7}{\sqrt{2\,7}} = \frac{2{,}23}{\sqrt{14}} = \frac{2{,}23}{3{,}74} = 0{,}59$$

Если $t_{\chi^2} < 3$, то расхождение между эмпирическим и теоретическим рядами случайно, что и оказалось в нашем примере

Разберем случай, когда требуется определить расхождение эмпирических и теоретических частот для качественных признаков, что довольно часто может встречаться в генетических работах

Пример. Сопоставить теоретическое и эмпирическое расщепления в соотношениях особей с доминантной и рецессивной окрасками у потомства второго поколения помесей F_2, полученных при скрещивании помесных родительских форм F_1, происходящих от коричневых мышей, подвергнутых облучению лучами рентгена табл 25 (см на стр 144) (p_1 и p_2 — соответственно частоты по окраске шерсти у мышей контрольной и подопытной групп)

$$\chi^2 = \frac{(65-75)^2}{75} + \frac{(35-25)^2}{25} = \frac{(-10)^2}{75} + \frac{(10)^2}{25} =$$

$$= 1{,}33 + 4 = 5{,}33,$$

Для того, чтобы оценить значение $\chi^2 = 5{,}33$, пользуемся той же табл 23 на стр 140

В указанном примере число степеней свободы ν равно числу классов, по варьирующему признаку без 1, т. е. $\nu = 2 - 1 = 1$.

По таблице при значениях $\nu = 1$ и $P = 0{,}95$ значение χ^2 должно быть 3,84; при $P = 0{,}99$ оно должно быть равным 6,63.

Таблица 25

Определение величины χ^2 для отклонения эмпирических частот от теоретических при качественном признаке

Окраска шерсти у мышей	Распределение мышей по окраске шерсти	
	теоретическое F_1	эмпирическое F_1
Коричневая	75 (p_1)	65 (p_2)
Белая	25 (p_1)	35 (p_2)

Полученное же значение χ^2 в разбираемом примере равно 5,33, т. е. оно меньше достоверности при $P = 0{,}99$. Это означает, что при такой требовательности к вероятности расщепление, полученное в опыте на облученных мышах, несущественно отличается от теоретического расщепления во втором поколении при моногибридном скрещивании, и, следовательно, примененное облучение исходных форм мышей не изменило наследования такого альтернативного признака, как окраска шерсти у потомства второго поколения.

Если число степеней свободы ν превышает 30, то для оценки χ^2 пользуются вычислением критерия его достоверности по следующей формуле:

$$t_{\chi^2} = \sqrt{2\chi^2} \cdot \sqrt{2\nu - 1},$$

причем, если $t_{\chi^2} > 2$, то разница между частотами существенна.

Если требуется произвести сопоставление эмпирических частот с теоретическими при малом числе наблюдений для альтернативных признаков, то обработка сохраняет указанный выше принцип.

Недостатки метода хи-квадрат заключаются в том, что 1) этот показатель зависит от системы перегруппировки материала и 2) он оценивает ряд разностей без учета порядка их следования, т. е. без учета их смысла по существу.

Недостаток метода хи-квадрат заключается также и в том, что с его помощью доказывается существенность разности, но не доказывается обратное, а именно, нельзя доказать «несущественность» разности.

4. ВЫРАВНИВАНИЕ ДАННЫХ МЕТОДОМ СКОЛЬЗЯЩЕЙ СРЕДНЕЙ

При построении эмпирических кривых изменения признака во времени, а также и для другого типа кривых, иногда наблюдаются резкие отклонения некоторых значений показателя от рядом стоящих с ним значений, что создает впечатление какой-то закономерности этих отклонений. На самом же деле это может быть вызвано случайностью и недостаточным числом наблюдений. Поэтому для выявления общей тенденции и закономерности в характере изменений показателя и выяснения случайности в таких отклонениях следует производить так называемое выравнивание рядов или отдельных его вариантов. Для этого пользуются методом скользящей средней или методом наименьших квадратов. Разберем один из примеров выравнивания, приведенный в книге А. Я. Боярского «Статистические методы в экспериментальных медицинских исследованиях», которая может служить также хорошим пособием по вопросам правильного использования статистики в биологических работах.

а. Метод скользящей средней. Предположим, что имеется следующий ряд изменения числа эритроцитов в крови крыс по дням опыта под влиянием введенного препарата:

Дни опыта	1	2	3	4	5	6	7	8	9	10	11
Число эритроцитов в *тыс*	5210	5100	5360	4590	5900	5120	5310	6030	5820	6640	6260

Если изобразить графически эту зависимость, то получаем ломаную кривую со спадом числа эритроцитов на 4-й и 9-й дни и увеличением их на 5-й и 10-й дни. Случайны ли эти колебания или они закономерны?

Методом скользящей средней можно выявить истинную тенденцию изменения численности эритроцитов под действием препарата. Этот метод заключается в том, что каждое значение изучаемого признака заменяется средней, получаемой из значения этого и соседнего с ним показателя по ряду. Для этого нужно вычислить попарные средние, затем попарные средние из этих средних и т. д., пока не получится результат, позволяющий заметить тенденцию в изменчивости данного признака в ряду.

В разбираемом примере нужно сделать так:

$$+\begin{array}{r}5210\\5100\\\hline 10310\end{array}\qquad +\begin{array}{r}5100\\5360\\\hline 10460\end{array}\qquad +\begin{array}{r}5360\\4590\\\hline 9950\end{array}\quad \text{и т. д.}$$

$$10310 : 2 = 5155 \qquad 10460 : 2 = 5230 \qquad 9950 : 2 = 4975$$

Таким путем получается ряд среднего числа эритроцитов по дням опыта, т. е. 5155; 5230; 4975; 5245; 5510, 5215; 5670; 5925; 6230; 6450.

Строим второй ряд попарных средних, т. е.

$$+\begin{array}{r}5155\\5230\\\hline 10385\end{array}\qquad +\begin{array}{r}5230\\4975\\\hline 10205\end{array}\quad \text{и т. д.}$$

$$10385 : 2 = 5192 \qquad 10205 : 2 = 5102$$

Получаем второй ряд попарных средних: 5192 (2-й день); 5102 (3-й день); 5110 (4-й день); 5377 (5-й день); 5362 (6-й день); 5442 (7-й день); 5797 (8-й день); 6077 (9-й день); 6340 (10-й день)[10].

---

[10] Примеры использования метода наименьших квадратов см. в книге А. Я. Боярский. Статистические методы в экспериментальных медицинских исследованиях. М., Медгиз, 1955.

ГЛАВА VIII

ДИСПЕРСИОННЫЙ АНАЛИЗ

1. ОСНОВНЫЕ ПОЛОЖЕНИЯ ДИСПЕРСИОННОГО АНАЛИЗА

Дисперсионный анализ — один из важнейших методов статистического анализа материала, который в значительной мере упрощает принцип составления выборочной совокупности.

Если в предыдущих главах разбирались методы обработки однородной по своему составу выборки при достаточном числе наблюдений, то при использовании дисперсионного метода в выборке может быть представлен качественно разнородный, структурно оформленный статистический комплекс как при большом, так и при ограниченном числе наблюдений. Эти новые методические стороны характерные для дисперсионного анализа, существенно расширяют возможность использования статистики для исследований биологических процессов, что тем более важно, так как структура и схемы физиологических, генетических и других опытов своеобразны. Они часто связаны с необходимостью вскрывать закономерности между изменчивостью изучаемых объектов при воздействии многих факторов, выяснение влияния которых раздельно в комплексе и ставится как цель исследования. Кроме того, для физиологических, а часто и генетических опытов, характерно включение в исследование незначительного числа объектов из-за сложности методики исследования или трудности получить в массовом количестве нужные для наблюдения объекты.

Малочисленность групп, подвергаемых изучению, часто сопутствует научным исследованиям в биологической науке. В то же время получаемые цифровые данные в подобного рода исследованиях требуют достоверности при достаточно высокой статистической вероятности для того, чтобы они могли служить характеристикой не только изученных в опыте объектов, а быть характеристикой генеральной совокупности в изучаемом явлении.

Биологические явления и процессы обладают той особенностью, что ход этих процессов обусловлен воздействием многих, часто противоречивых по своему действию, факторов. Выявление доли влияния на интересующий нас показатель или процесс каждого из действующих факторов и их комплексного воздействия на изменчивость признака и составляет задачу биологической науки. Но решение этой задачи специальными приемами и методами исследования должно сопровождаться математической обработкой получаемых данных В этом случае лучше всего поможет *метод дисперсионного анализа, разработанный впервые* Р Фишером в 1925 г. и продолженный другими исследователями [11]

В задачу дисперсионного анализа входит определение доли влияния различных факторов в отдельности, а также и суммарного их воздействия на изменчивость данного признака

В ходе этого анализа получаются данные, характеризующие *общую дисперсию* (*изменчивость*) признака, обусловленную действием всевозможных учтенных и неучтенных факторов, частную (или факториальную) дисперсию, т е ту долю изменчивости, которая обусловлена воздействием различных факторов, учтенных исследователем, остаточную дисперсию, т. е изменчивость, вызванную неизвестными исследователю причинами

Если до сих пор в предыдущих главах излагались методы, позволяющие измерить изменчивость в целом с помощью лимитов, среднего квадратического отклонения или коэффициента изменчивости, то в дисперсионном анализе степень изменчивости определяется сопоставлением показателя каж-

---

[11] *Для ознакомления с этим методом можно использовать следующие* наиболее доступные по изложению книги

Н А Плохинский Дисперсионный анализ Новосибирск, Изд-во Сиб отделения АН СССР, 1960,

А К Митропольский Элементы статистического исчисления Л, ВЗЛТИ, 1957,

Ю Л Поморский Новейшие методы вариационной статистики Л, 1939,

В Н Перегудов Статистические методы обработки данных полевого опыта М, Сельхозгиз, 1948,

Сб «Методика полевого опыта» М, Сельхозгиз, 1959,

В И Романовский Применения математической статистики в опытном деле М — Л, Гостехиздат, 1947.

дой особи выборки со средней арифметической, возведением этих отклонений в квадрат и дальнейшим их суммированием.

В общей форме это может быть выражено следующим образом:

$$\Sigma(V-M)^2 \quad \text{или} \quad C=\Sigma D^2=\Sigma(V-M)^2\,[12].$$

Часто дисперсия по какому-либо из факторов выражается суммой квадратов отклонений частной средней от общей средней арифметической:

$$C=\Sigma(M_{\text{частн}}-M_{\text{общ}})^2.$$

Если обозначить через C величину общей дисперсии, через y — результативный признак, через x — организованные факторы воздействия (т. е. контролируемые исследователем) и через z — неорганизованные факторы, остающиеся вне контроля исследователя, то общая изменчивость или общая дисперсия признака y может быть представлена в принципе таким выражением:

$$C_y=C_x+C_z,$$

т. е. сумма факториальной и остаточной дисперсий будет давать значение общей дисперсии результативного признака.

Если учтенных факторов несколько, например, два (A — температура и B — влажность воздуха), то графически можно показать, что результативный признак (например, урожайность) y будет находиться под воздействием следующих факторов: A, B, их совместного действия AB и z (рис. 21).

Следовательно, при проведении дисперсионного анализа предстояло бы выяснить изменчивость результативного признака y (урожайности) под влиянием каждого фактора в их раздельном действии, т. е. найти величину частных дисперсий C_A, C_B, а также величину изменчивости, вызываемой совместным действием обоих факторов, т. е. C_{AB}. Кроме того, необходимо было бы найти и ту долю изменчивости, которая возникает под влиянием прочих неучтенных факторов, т. е. определить C_z. Таким образом, в этом случае общая дисперсия $C_{\text{общ}}$ может быть представлена выражением:

$$Cy=C_A+C_B+C_{AB}+C_z.$$

---

[12] В некоторых книгах дисперсия обозначается через S.

Следует иметь в виду, что чем выше доля изменчивости обусловлена изучаемыми факторами и чем меньше доля остаточной (случайной) дисперсии C_z, тем больше может быть познана изменчивость изучаемого нами объекта при помощи дисперсионного анализа.

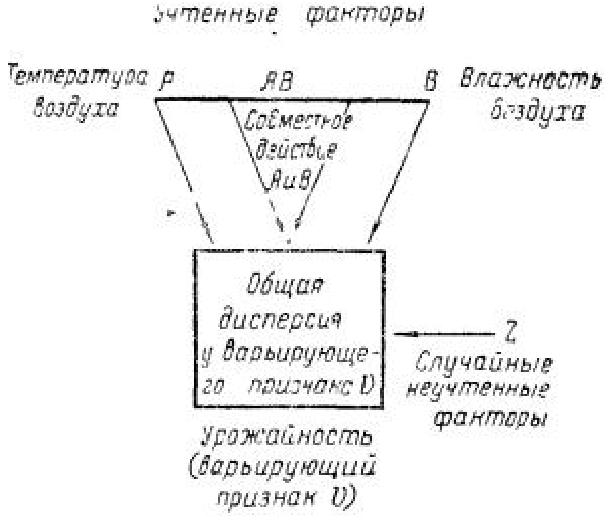

Рис. 21

Дисперсионный анализ складывается из трех этапов.

Первый этап заключается в отыскании значений C_y, C_x и C_z, т. е. общей, факториальной и остаточной дисперсии.

Если общая дисперсия C_y может быть выражена суммой квадратов отклонения каждого варианта от общей средней, т. е. $C_y = \Sigma (v - M_{общ})^2$, то факториальная дисперсия C_x, т. е. изменчивость, обусловленная влиянием учтенных факторов, выражается суммой квадратов отклонения частных средних арифметических от общей средней арифметической:

$$C_x = \Sigma n_x (M_{частн} - M_{общ})^2,$$

где n указывает число наблюдений по факторам.

Факториальную дисперсию C_x можно выразить и так:

$$C_x = \Sigma n_x \cdot D_x^2.$$

Случайная, или остаточная дисперсия C_z выражается суммой квадратов разницы между каждым вариантом, вошедшим в таблицу, и частной средней арифметической:

$$C_z = \Sigma (V - M_x)^2 = \Sigma D_z^2.$$

Второй этап дисперсионного анализа заключается в отнесении каждой из дисперсий C к числу степеней свободы ν. Этот прием вызван тем, что степень варьирования признака V зависит от числа степеней свободы, которое в свою очередь связано с числом наблюдений и количеством ограничивающих условий.

В самом простом виде смысл числа степеней свободы вытекает из такого примера: если имеется вариационный ряд из пяти наблюдений со значениями варьирующего признака V, равных 10, 8, 11, 9, 13, то число степеней свободы для него будет равно $\nu = n - 1 = 5 - 1 = 4$. Это означает, что из числа данных наблюдений имеется одно ограничение. Таким ограничением в статистическом ряде является сумма всех вариантов. Это значит, что если сумма вариантов $\Sigma V = 10 + 8 + 11 + 9 + 13 = 51$, то для получения этой суммы можно свободно менять значения V_y четырех членов ряда, а значение пятого члена ограничивается разницей между суммой пяти и четырех членов.

Так, предположим, что при $\Sigma V = 51$ четыре члена ряда имеют следующие значения: 6, 5, 15, 18. Тогда на пятый член остается значение, равное $51 - 44 = 7$.

С числом степеней свободы пришлось встретиться при рассмотрении метода хи-квадрат и увидеть, что оно состоит из числа наблюдений или классов, уменьшенных на число условий, ограничивающих свободу варьирования. В последнем примере ограничивающим условием было значение $\Sigma V = 51$.

При дисперсионном анализе число степеней свободы зависит от характера таблицы статистического комплекса, подвергнутого обработке, от числа рядов и столбцов в таблице. В дальнейшем будет показано, как следует определять число степеней свободы для общей, факториальной и остаточной дисперсий.

Отношение общей и частной дисперсий к числу степеней свободы ν позволяет, таким образом, связать изменчивость со структурой статистического комплекса.

Принято обозначать частное, полученное от деления дисперсии C на число степеней свободы ν, через σ^2. Часто этот показатель называют девиатой. Но следует иметь в виду, что во многих пособиях это выражение называется и величиной дисперсии. В дальнейшем будем и значение C, и значение σ^2 называть **дисперсией**.

Итак, дисперсии, корректированные на число степеней свободы, будут иметь следующие выражения:
общая дисперсия

$$\sigma_y^2 = \frac{C_y}{\nu_y};$$

частная, или факториальная дисперсия

$$\sigma_x^2 = \frac{C_x}{\nu_x};$$

остаточная, или случайная дисперсия

$$\sigma_z^2 = \frac{C_z}{\nu_z}.$$

Уместно заметить, что выражение $\sigma = \sqrt{\frac{C}{\nu}}$ дает известное уже нам значение среднего квадратического отклонения, в которое входят квадраты отклонений вариантов от средней арифметической.

В расшифрованном виде дисперсии можно представить и так:

$$\sigma_y^2 = \frac{\Sigma (V - M_{\text{общ}})^2}{\nu_y};$$

$$\sigma_x^2 = \frac{\Sigma n_x (M_x - M_{\text{общ}})^2}{\nu_x};$$

$$\sigma_z^2 = \frac{\Sigma (V - M_x)^2}{\nu_z}.$$

Третий этап дисперсионного анализа сводится к выяснению достоверности влияния каждого фактора. Для этого значение факториальной (частной) дисперсии относится к остаточной дисперсии. Это отношение обозначается буквой F (по начальной букве фамилии автора метода—Фишера) и служит показателем достоверности.

Достоверность факториальных дисперсий приобретает, таким образом, следующее выражение:

$$F = \frac{\sigma_x^2}{\sigma_z^2}.$$

Для суждения о достоверности влияния того или иного учтенного при анализе фактора пользуются таблицами Фишера для значений F (см. «Приложение»). При этом значение F берется с учетом числа степеней свободы у σ_x^2 и σ_z^2. В зависимости от того, на каком уровне вероятности хотят иметь заключение о достоверности, берут в таблице значение F или при низком уровне вероятности ($P_1 = 0{,}95$), или при среднем значении ($P_2 = 0{,}99$), или при высоком ($P_3 = 0{,}999$). Таблица составлена так, что верхняя заглавная строка и первый столбец означают соответственно число степеней свободы ν_1 и ν_2, а на пересечении их стоят три значения коэффициента F (при трех градациях вероятности P).

Если вычисленное по нашим данным значение $F = \frac{\sigma_x^2}{\sigma_z^2}$ больше табличного значения F при заданных величинах ν_x и ν_z и заданном уровне P, то расхождение факториальной дисперсии от остаточной существенно, что указывает на достоверность влияния изучаемых факторов на результативный признак.

При обработке материала дисперсионным анализом можно получить выражение доли влияния того или иного фактора в общей дисперсии признака, которую можно принять за единицу или 100%. Для этого факториальную и остаточную дисперсии нужно отнести к общей дисперсии; это отношение выражается η^2, а именно:

$$\eta_x^2 = \frac{C_x}{C_y} \text{ — доля влияния учтенных факторов;}$$

$$\eta_z^2 = \frac{C_z}{C_y} \text{ — доля влияния случайных факторов.}$$

Если взять корень квадратный из первого выражения, то получим формулу корреляционного отношения, о чем говорилось ранее на стр. 101.

Таким образом, $\eta_{x/y}^2 = \frac{C_x}{C_y}$ есть мера влияния организованного фактора на общую изменчивость признака. Если общую изменчивость принять за единицу, то доли ее будут приходиться на величину влияния каждого фактора и на случайные воздействия, т. е.

$$\eta_y^2 = \eta_x^2 + \eta_z^2 = 1 \text{ (или 100\%).}$$

Изложенное выше о принципах дисперсионного анализа позволяет представить значение этого метода в биологических исследованиях.

Р. Фишер разработал для полевых опытов такие приемы, при которых, пользуясь дисперсионным анализом, можно установить доли изменчивости признака от факторов и устранить влияние почвенных различий на результативный признак (урожайность). Это — приемы так называемых случайных блоков и латинского квадрата. Они хорошо разобраны в книге В. И. Романовского, которую и рекомендуется использовать для руководства при закладке и обработке полевых опытов.

Приемы дисперсионного метода теперь используются довольно широко в генетических исследованиях. Так, доля генотипической изменчивости, обусловленная наследственностью, вычленяется с использованием дисперсионного анализа. Этот подход применяется зарубежными генетиками и селекционерами при оценке производителей по качеству потомства. При этом также вычленяется доля ненаследственной изменчивости. Не вдаваясь в критику, насколько целесообразно и насколько безупречно можно пользоваться дисперсионным анализом для некоторых (в частности, для генетических исследований), следует отметить общие недостатки в самой методике дисперсионного анализа, что хорошо подчеркнуто в упомянутой книге А. Я. Боярского (см. сноску на стр. 146).

Недостаток дисперсионного анализа состоит в том, что результат оценки по факторам может зависеть от системы перегруппировки материала, т. е. от того, как составлена рабочая таблица статистического комплекса, по которой ведут вычисление. Так, например, если между средними значениями по группам имеется резкая разница, а в подгруппах средние показатели близки, то дисперсионный анализ может привести к выводу, отрицательному в отношении наличия существенной разницы между средними.

Следует иметь в виду, что если в дисперсионном анализе получен вывод о несущественности разницы, то это не служит основанием для суждения, так как дисперсионный анализ всегда дает правильное суждение только для случаев существенной разницы.

А. Я. Боярский в книге «Статистические методы в экспериментальных медицинских исследованиях» подчеркивает, что иногда неравнозначность вывода о наличии несущественной разницы представителями моргановской генетики отождествлялась с выводом об отсутствии существенной разницы в воздействиях на признак; это в статистическом понимании совершенно не равнозначно по своему математическому смыслу, т. е. представители этого направления генетики превращали недоказанность «существенности» в доказанность несущественности связи между действующими факторами и варьирующим признаком организма.

Несмотря на указанные недостатки, метод дисперсионного анализа значительно расширяет возможности научной оценки результатов опыта.

Перейдем к разбору примеров и техники вычисления показателей дисперсии.

Выше уже отмечалось, что дисперсионный анализ осуществляется на структурно оформленном статистическом комплексе, а не на вариационном ряду, с которыми приходится иметь дело во всех предыдущих обработках материала. От правильного подбора объектов, показатели которых служат результативным признаком, и от правильного построения таблицы статистического комплекса зависит правильность дисперсионного анализа.

Прежде всего обязательно условие, что в статистический комплекс включаются особи по принципу случайной выборки, которая была обязательна и для всех предыдущих методов статистической обработки. Применение случайного отбора объектов называется рэндомизацией (т. е. выборкой наугад), а составленный из таких объектов статистический комплекс — рэндомизированным.

Структура статистического комплекса обусловлена тем, какие факторы и в каком количестве изучаются дисперсионным методом.

Так, если нас интересует такой результативный признак, как плодовитость мышей, то можно изучать действие одного фактора, например, облучения, разными дозами. Число изучаемых доз даст несколько градаций (или классов) этого фактора. В этом случае статистический комплекс будет однофакторным. Если же плодовитость мышей нас интересует не только в связи с облучением, но и в связи с полом облучаемых животных, то тогда статистический комплекс усложняется, превращается в двухфакторный (облучение и пол). Могут быть и более сложные, т. е. трех- и четырехфакторные статистические комплексы. При этом необходимо условие, чтобы воздействующие организованные факторы не находились между собой в коррелятивной связи.

Кроме числа изучаемых факторов статистические комплексы отличаются еще и размещением членов комплекса по градациям (или классам) комплекса. В этом отношении комплексы могут быть *равномерными, пропорциональными* и *неравномерными*.

В равномерных комплексах число наблюдений (частоты) в каждом классе одного фактора равно числу частот по градациям другого фактора.

В пропорциональных комплексах в разных градациях (классах) одного фактора (A) частоты другого фактора (B) имеют равные отношения; в неравномерных комплексах в разных классах частоты другого комплекса таким свойством не обладают.

Однофакторные комплексы всегда относятся к типу пропорциональных, независимо от отношений частот по классам.

В качестве примера разберем построение этих комплексов по показателю плодовитости самок мышей в зависимости от дозы облучения (фактор A) и пола (фактор B) облученных животных (табл. 26).

Предположим, что облучение изучается в таких градациях: 0 рентген (контроль), 100 рентген и 200 рентген. Облучение проводилось на одновозрастных самцах и самках.

Таким образом, здесь имеет место двухфакторный статистический комплекс, в котором по фактору A имеем три класса (A_1, A_2 и A_3) и по фактору B — два класса (т. е. B_1 — самцы и B_2 — самки). Отсюда ясно, что в данном статистическом комплексе будет 6 клеток ($3 \cdot 2$), обусловленных числом классов обоих факторов. В этих 6 клетках разместятся показатели всех подопытных и контрольных животных, число которых, положим, равно 12.

Предположим, что комплекс составлен был при равномерном соотношении частот. Это означает, что в каждой клетке таблиц будут проставлены сведения о плодовитости двух животных, а сам статистический комплекс будет выглядеть следующим образом (табл. 26).

Таблица 26

Равномерный статистический комплекс

	$A_1=0$		$A_2=100$		$A_3=200$	
	$B_1=♂$	$B_2=♀$	$B_1=♂$	$B_2=♀$	$B_1=♂$	$B_2=♀$
Варианты (плодовитость), V	10; 12	11; 10	8; 10	7; 9	7; 9	6; 4
Частоты, p	2	2	2	2	2	2

Таким образом, в каждой клетке этого равномерного комплекса было по два наблюдения, и отношение частот p по всем клеткам равно 1 · 1.

Если бы комплекс был не равномерный, а пропорциональный, то 12 объектов могли быть распределены по клеткам комплекса следующим образом (табл 27)

Таблица 27

Пропорциональный статистический комплекс

	$A_1=0$		$A=100$		$A_1=200$	
	$B_1=♂$	$B=♀$	$B_1=♂$	$B=♀$	$B_1=♂$	$B=♀$
Варианты (плодовитость), V	10	11, 10	8, 10	7, 9, 12, 9	7	6, 4
Частоты, p	1	2	2	4	1	2
Отношения частот	1 2		1 2		1 2	

Как видно, здесь отношения частот фактора B по всем классам фактора A равны 1 : 2.

Неравномерный комплекс для того же примера будет выглядеть так (табл 28)

Таблица 28

Неравномерный статистический комплекс

	$A_1=0$		A 100		$A_3=200$	
	$B_1-♂$	$B=♀$	$B_1=♂$	$B_2=♀$	$B_1-♂$	$B_2=♀$
Варианты (плодовитость), V	10	11, 10	8, 10	7, 9, 12	7, 9	6, 4
Частоты, p . .	1	2	2	3	2	2
Отношения частот	1 : 2		1 : 1,5		1 : 1	

Откуда видно, что градации фактора B имеют разные отношения по градациям фактора A.

Техника обработки статистических комплексов зависит от числа факторов комплекса, от наличия или отсутствия пропорциональности между частотами классов, а также от общего числа наблюдений в комплексе (малое или большое) [13].

2. ТЕХНИКА ОБРАБОТКИ СТАТИСТИЧЕСКОГО КОМПЛЕКСА ПРИ ДИСПЕРСИОННОМ АНАЛИЗЕ

При обработке статистических комплексов используются формулы, в основе которых лежат те общие формулы, которые разбирались в предыдущей главе.

А. Равномерные и пропорциональные комплексы

а. Обработка однофакторного комплекса при малом числе наблюдений — это наиболее простой случай применения дисперсионного анализа.

Разберем пример, в котором приведены данные, характеризующие изменение обхвата пясти у коров джерсейской породы трех эколого-генетических генераций (табл. 29).

Из табл. 29 виден ход последовательной обработки статистического комплекса. Для вычисления общей, факториальной и остаточной дисперсий пользуются следующими формулами:

$$C_y = \Sigma V^2 - \frac{(\Sigma V)^2}{n} \quad \text{или} \quad C_y = \Sigma V^2 - H.$$

Подобное выражение уже встречалось на стр. 43, когда вычислялось значение α, входящее в формулу среднего квадратического отклонения σ.

По данным примера общая дисперсия будет следующей:

$$C_y = \Sigma V^2 - H = 1749 - 1736{,}1 = 12{,}9.$$

Вычислим остаточную дисперсию по следующей формуле:

$$C_z = \Sigma V^2 - \Sigma h = 1749 - 1745{,}6 = 3{,}4.$$

[13] Техника обработки статистических комплексов изложена в книге: Н. А Плохинский. Дисперсионный анализ. Новосибирск, Изд-во Сиб. отделения АН СССР, 1960.

Значение $\Sigma h_x = \Sigma \frac{(\Sigma V_x)^2}{n_x}$ также уже известно из приемов вычисления среднего квадратического отклонения.

Таблица 29

Обработка однофакторного комплекса при малом числе наблюдений

Число классов по фактору A $l = 3$	Первая генерация, A_1	Вторая генерация, A_2	Третья генерация, A_3	
Варьирующий признак (обхват пясти в *см*), V	14; 12; 13	14; 13; 13	15; 15; 16	$\Sigma V = 125$
V^2	196; 144; 169	196; 169; 169	225; 225; 256	$\Sigma V^2 = 1749$
n_x	3	3	3	$\Sigma n_x = 9$
ΣV_x	39	40	46	—
$(\Sigma V_x)^2$	$39^2 = 1521$	$40^2 = 1600$	$46^2 = 2116$	—
$h = \frac{(\Sigma V_x)^2}{n_x}$	$\frac{1521}{3} = 507,0$	$\frac{1600}{3} = 533,3$	$\frac{2116}{3} = 705,3$	$\Sigma h_x = 1745,6$
$M_x = \frac{\Sigma V_x}{n_x}$	$\frac{39}{3} = 13$	$\frac{40}{3} = 13,3$	$\frac{46}{3} = 15,3$	$M_{общ} = \frac{\Sigma V}{n} = \frac{125}{9} = 13,8$

$$H = \frac{(\Sigma V)^2}{n} = \frac{125^2}{9} = \frac{15625}{9} = 1736,1$$

Факториальная дисперсия вычисляется по формуле:

$$C_x = \Sigma h_x - H = 1745,6 - 1736,1 = 9,5.$$

Таким образом:

$$C_y = C_x + C_z = 9{,}5 + 3{,}4 = 12{,}9.$$

Доля влияния организованных факторов η_x^2 равна следующему значению:

$$\eta_x^2 = \frac{C_x}{C_y} \cdot 100 = \frac{9{,}5}{12{,}9} \cdot 100 = 73{,}6\%.$$

Доля влияния неучтенных факторов равна:

$$\eta_z^2 = \frac{C_z}{C_y} \cdot 100 = \frac{3{,}4}{12{,}9} \cdot 100 = 26{,}3\%.$$

Следовательно, доля влияния экологического фактора в увеличении крепости костяка коров джерсейской породы велика и составляет 73,6% общей изменчивости.

Дальше при помощи дисперсионного анализа определяем дисперсию, отнесенную к числу степеней свободы, и затем определяем достоверность влияния организованных факторов.

Обычно дисперсионный анализ осуществляется составлением сводной таблицы в виде следующей стандартной формы, которая заполняется конкретными величинами, получаемыми в ходе вычисления (табл. 30).

Таблица 30

Сводная таблица дисперсионного анализа

	x	z	y
Дисперсия, C	$\Sigma h_x - H$	$\Sigma V^2 - \Sigma h_x$	$\Sigma V^2 - H$
Степень влияния фактора, η^2	$C_x : C_y$	$C_z : C_y$	1,0
Число степеней свободы, ν	$l - 1$	$n - l$	$n - 1$
Взвешенная дисперсия (девиата), σ^2	$C_x : \nu_x$	$C_z : \nu_z$	—
Коэффициент достоверности, F . .	$\sigma_x^2 : \sigma_z^2$	—	—

Заполним такую стандартную сводную таблицу дисперсионного анализа данными из приведенного примера (табл. 31).

Для определения достоверности влияния фактора x сопоставим вычисленное значение $F = 8{,}4$ с табличным значением F. В таблице «Приложение» на стр. 222 значение числа степеней свободы ν_1 принимаем за факториальное число степеней свободы ν_x, т. е. $\nu_x = 2$, а значение ν_2 — за значение степеней свободы для остаточной дисперсии, т. е. за $\nu_z = 6$.

При $\nu_1 = 2$ и $\nu_2 = 6$ коэффициент достоверности F имеет следующие значения при трех градациях вероятности: $F = 27{,}0$ при $P = 0{,}999$; $F = 10{,}9$ при $P = 0{,}99$ и $F = 5{,}1$ при $P = 0{,}95$.

Таблица 31

Сводная таблица дисперсионного анализа (к примеру на стр. 159)

	x	z	y
Дисперсия, C	$\Sigma h_x - H = 9{,}5$	$\Sigma V^2 - \Sigma h_x = 3{,}4$	$\Sigma V^2 - H = 12{,}9$
Влияние факторов в %, η^2	$C_x : C_y = 73{,}6$	$C_z : C_y = 26{,}3$	100,0
Число степеней свободы, ν	$l-1=3-1=2$	$n-l=9-3=6$	$n-1=9-1=8$
Взвешенная дисперсия, σ^2	$\sigma_x^2 = \frac{9{,}5}{2} = 4{,}75$	$\sigma_z^2 = \frac{3{,}4}{6} = 0{,}56$	—
Коэффициент достоверности, F	$\frac{\sigma_x^2}{\sigma_z^2} = \frac{4{,}75}{0{,}56} = 8{,}4$	—	—

Так как вычисленное значение $F = 8{,}4$ больше табличного только при пониженном уровне вероятности $P = 0{,}95$, то достоверность влияния экологического фактора имеет место при пониженном требовании к достоверности, которая может быть использована в такого рода наблюдениях, где не требуется более высокого уровня достоверности.

б. Обработка однофакторного комплекса при большом числе наблюдений. При большом числе наблюдений для дисперсионного анализа удобно пользоваться обычной корреляционной решёткой, в которой один из показателей является зависимым (результативный признак y), а другой соответствует фактору воздействия x.

Обработка такой решётки осуществляется комбинированным методом, в котором сочетаются приемы, используемые при вычислении коэффициента корреляции и корреляционного отношения.

Разбивка на классы и разноска данных по клеткам корреляционной решётки осуществляется обычным способом, что изложено на стр. 81, 82.

Разберем на примере ход дисперсионного анализа при большом числе наблюдений.

Пример. Определить долю влияния длительности акклиматизации в новых экологических условиях на жирномолочность коров джерсейской породы (табл. 32).

Таблица 32

Однофакторный комплекс при большом числе наблюдений

Жирномолочность, y %	Год акклиматизации, x: 1	2	3	4	5	p_y	a_y	$p_y \cdot a_y$	$p_y \cdot a_y^2$
5,0	1	—	—	—	—	1	—4	—4	16
5,25	2	1	1	—	—	4	—3	—12	36
5,50	3	3	6	3	—	15	—2	—30	60
5,75	5	7	10	5	—	27	—1	—27	27
6,00	7	13	8	2	—	30	0	0	0
6,25	2	1	3	2	2	10	1	10	10
6,50	—	—	1	2	3	6	2	12	24
6,75	—	—	1	1	3	5	3	15	45
7,00	—	—	—	—	2	2	4	8	32
p_x	20	25	30	15	10	100	—	$S_1 = \Sigma p_y \cdot a_y = -28$	$S_2 = \Sigma p_y \cdot a_y^2 = 250$
$\Sigma p_{xy} \cdot a_y$	—19	—15	—17	—2	25	$\Sigma p_{xy} \cdot a_y = -28$	—	—	—
$(\Sigma p_{xy} \cdot a_y)^2$	361	225	289	4	625	—	—	—	—
$h_x = \dfrac{\Sigma(p_{xy} \cdot a_y)^2}{p_x}$	$\dfrac{361}{20} = 18{,}05$	$\dfrac{225}{25} = 9{,}0$	$\dfrac{289}{30} = 9{,}63$	$\dfrac{4}{15} = 0{,}26$	$\dfrac{625}{10} = 62{,}5$	$\Sigma h_x = 99{,}44$			

Для того, чтобы яснее был ход обработки таблицы, приведем рабочие формулы дисперсионного анализа для большой однофакторной выборки.

Общая дисперсия: $C_y = S_2 - H$.

Здесь $S_2 = \Sigma p_y \cdot a_y^2$ получается от обработки ряда y методом произведений, т. е. умножением частот ряда y на соответствующее условное отклонение a_y или методом сумм, тогда $S_2 = q_2 + q_1 + 2(q_4 + q_3)$, где q_1, q_2, q_3, q_4 представляют суммы накопленных частот двух столбиков, полученных при обработке ряда методом сумм (см. стр. 41—42).

Значение H вычисляется следующим образом: $H = \frac{S_1^2}{n}$, что может быть представлено и так: $H = \frac{(\Sigma p_y \cdot a_y)^2}{n}$. По способу сумм значение S_1 было бы равно $q_2 - q_1$ и давало бы то же значение, что получается для величины S_1, вычисленной по методу произведений.

Факториальная дисперсия: $C_x = \Sigma h_x - H$.

Получение для этой формулы величины Σh_x осуществляется приемом, который был использован при вычислении корреляционного отношения, что видно из формулы:

$$\Sigma h_x = \Sigma \left\{ \frac{(p_{xy} \cdot a_y)^2}{p_x} \right\}.$$

Система обработки этого выражения изложена на стр. 106. Из самой формулы видно, что необходимо составить строку из значений $p_{xy} \cdot a_y$ при помощи умножения по столбцам частот клеток таблицы на соответствующее значение условного отклонения по признаку y, т. е. на a_y. Суммирование этих данных по столбцам решётки дает строчку, обозначаемую $\Sigma p_{xy} \cdot a_y$. Затем получается строчка $(\Sigma p_{xy} \cdot a_y)^2$ от возведения в квадрат данных предыдущей строчки по каждому столбцу. Последнее действие при обработке решётки заключается в получении строчки, выражаемой как $(\Sigma p_{xy} \cdot a_y)^2 : p_x$. Подобные значения по столбцам получаются от деления данных предыдущей строчки по столбцам на соответствующее значение частот ряда x, т. е. на p_x. Суммируя полученные частные значения по всей строчке, получаем значение Σh_x, входящее в формулу факториальной дисперсии.

Величина $H = \frac{(\Sigma p_y \cdot a_y)^2}{n}$ уже известна, так как вычислена для формулы общей дисперсии.

Остаточная дисперсия: $C_z = S_2 - \Sigma h_x$.

Здесь значения S_2 и Σh_x уже известны.

Имея в виду эти разъяснения, произведем обработку табл. 32 по признаку y методом произведений для получения $\Sigma p_y \cdot a_y$ и $\Sigma p_y \cdot a_y^2$ и обработку ряда x для получения $\Sigma h_x = \{(\Sigma p_{x\,y} \cdot a_y)^2 : p_x\}$. В результате чего $S_1 = \Sigma p_y \cdot a_y = -28$, а $S_2 = \Sigma p_y \cdot a_y^2 = 250$.

Зная S_1, находим:

$$H = \frac{S_1^2}{n} = \frac{(-28)^2}{100} = \frac{784}{100} = 7{,}84.$$

Для получения h_x делаем последовательные вычисления.

Находим $\Sigma p_{xy} \cdot a_y$:

1-й столбец: $1 \cdot (-4) = -4$; $2 \cdot (-3) = -6$; $3 \cdot (-2) = -6$; $5 \cdot (-1) = -5$; $7 \cdot 0 = 0$; $2 \cdot 1 = 2$; $\Sigma p_{xy} \cdot a_y = -19$.

2-й столбец: $1 \cdot (-3) = -3$; $3 \cdot (-2) = -6$; $7 \cdot (-1) = -7$; $13 \cdot 0 = 0$; $1 \cdot 1 = 1$; $\Sigma p_{xy} \cdot a_y = -15$.

3-й столбец: $1 \cdot (-3) = -3$; $6 \cdot (-2) = -12$; $10 \cdot (-1) = -10$; $8 \cdot 0 = 0$; $3 \cdot 1 = 3$; $1 \cdot 2 = 2$; $1 \cdot 3 = 3$; $\Sigma p_{xy} \cdot a_y = -17$.

4-й столбец: $3 \cdot (-2) = -6$; $5 \cdot (-1) = -5$; $2 \cdot 0 = 0$; $2 \cdot 1 = 2$; $2 \cdot 2 = 4$; $1 \cdot 3 = 3$; $\Sigma p_{xy} \cdot a_y = -2$.

5-й столбец: $2 \cdot 1 = 2$; $3 \cdot 2 = 6$; $3 \cdot 3 = 9$; $2 \cdot 4 = 8$; $\Sigma p_{xy} \cdot a_y = 25$.

По всем столбцам $\Sigma p_{xy} \cdot a_y = -19 - 15 - 17 - 2 + 25 = -28$.

Значение $\Sigma p_{xy} \cdot a_y$ всегда должно быть равно значению $\Sigma p_y \cdot a_y$, что служит проверкой правильности вычисления по этим столбцам.

Далее возводим в квадрат значение $\Sigma p_{xy} \cdot a_y$, полученное в каждом столбце: $(-19)^2 = 361$; $(-15)^2 = 225$; $(-17)^2 = 289$; $(-2)^2 = 4$; $25^2 = 625$.

Эти величины записываем в строчку, обозначенную $(\Sigma p_{xy} \cdot a_y)^2$.

Далее, полученное в каждом столбце значение $(\Sigma p_{xy} \cdot a_y)^2$, делим на соответствующее значение частот ряда x, т. е. на p_x и суммируем полученные дроби. Получаем значения:

$$\frac{361}{20} = 18{,}05; \quad \frac{225}{25} = 9; \quad \frac{289}{30} = 9{,}63; \quad \frac{4}{15} = 0{,}26; \quad \frac{625}{10} = 62{,}5;$$

откуда: $\Sigma h_x = 18{,}05 + 9 + 9{,}63 + 0{,}26 + 62{,}5 = 99{,}44$.

Все необходимые параметры для определения общей, частной и случайной дисперсий найдены, вычислим их значения:

$$C_y = S_2 - H = 250 - 7{,}84 = 242{,}16;$$

$$C_x = \Sigma h_x - H = 99{,}44 - 7{,}84 = 91{,}60;$$

$$C_z = S_2 - \Sigma h_x = 250 - 99{,}44 = 150{,}56.$$

Находим степень влияния каждого фактора на изучаемый признак:

$$\eta_x^2 = \frac{C_x}{C_y} = \frac{91{,}60}{242{,}16} = 0{,}378,$$

т. е. из общей изменчивости 37,8% составляет влияние на процент жира в молоке, обусловленное длительностью акклиматизации;

$$\eta_z^2 = \frac{C_z}{C_y} = \frac{150{,}56}{242{,}16} = 0{,}621,$$

т. е. 62,1% влияния приходится на прочие, неорганизованные факторы.

Попутно можно вычислить и корреляционное отношение, которое показывает величину связи между воздействием длительности акклиматизации и жирномолочностью. На стр. 101 уже указывалось, что корреляционное отношение может быть вычислено по формуле:

$$\eta_{xy} = \sqrt{\frac{C_x}{C_y}} = \sqrt{\frac{91{,}60}{242{,}16}} = \sqrt{0{,}378} = 0{,}614,$$

т. е. связь между указанными признаками высокая. Статистическая ошибка корреляционного отношения равна:

$$m_\eta = \frac{1 - \eta^2}{\sqrt{n}} = \frac{1 - 0{,}378}{\sqrt{100}} = \frac{0{,}622}{10} = 0{,}0622.$$

Достоверность этой связи велика и равна

$$t_\eta = \frac{\eta}{m_\eta} = \frac{0{,}614}{0{,}0622} \approx 10.$$

Продолжим далее ход дисперсионного анализа.

Определим σ^2, т. е. дисперсию, отнесенную к числу степеней свободы, а также определим достоверность доли влияния фактора x на результативный признак y. Для указанного требуется найти число степеней свободы для y, x и z. Их значения будут вытекать из следующих формул:

для общей дисперсии $\nu_y = n-1$, т. е. число наблюдений, уменьшенное на единицу $(\nu_y = 100 - 1 = 99)$;

для факториальной дисперсии $\nu_x = l-1$, т. е. число классов l по факториальному признаку x, уменьшенное на единицу $(\nu_z = 5 - 1 = 4)$;

для остаточной дисперсии $\nu_z = n - l$, т. е. число наблюдений, уменьшенное на число классов по признаку x.

Следовательно, $\nu_z = 100 - 5 = 95$.

Полезно использовать правило в качестве контроля за верностью определения значения ν, а именно: сумма частных степеней свободы должна быть равна числу степеней свободы для общей дисперсии, т. е. $\nu_x + \nu_z = \nu_y$. В нашем примере это даст: $4 + 95 = 99$.

Находим значения:

$$\sigma_x^2 = \frac{C_x}{\nu_x} = \frac{91,60}{4} = 22,9;$$

$$\sigma_z^2 = \frac{C_z}{\nu_z} = \frac{150,56}{95} = 1,58.$$

Определяем критерий достоверности влияния фактора x на результативный признак y, т. е. находим значение F:

$$F \text{ (расчетное)} = \frac{\sigma_x^2}{\sigma_z^2} = \frac{22,9}{1,58} = 14,49.$$

По таблице Фишера, приведенной в «Приложении» (см. стр. 222), значения степеней свободы при $\nu_1 = \nu_x = 4$ и $\nu_2 = \nu_z = 95$; при таких значениях ν табличное значение F при вероятности $P = 0,999$ равно 5,0; при $P = 0,99$ равно 3,5 и при $P = 0,95$ оно равно 2,5.

Так как вычисленное значение $F = 14,49$ больше табличного значения F при всех градациях вероятности P, то можно утверждать, что доля влияния изучаемого фактора на результативный признак, составляющая 38,0% общей изменчивости, вполне достоверна.

Дисперсионный анализ оформляется в виде следующей итоговой таблицы 33.

Сводный результат дисперсионного анализа при однофакторном комплексе и большом числе наблюдений

	Организованные факторы, x	Неорганизованные факторы, z	Общая изменчивость, y
Дисперсия, C	$\Sigma h_x - H = 91,60$	$S_2 - \Sigma h_x = 150,56$	$S_2 - H = 242,16$
Степень влияния факторов, η^2	$\frac{C_x}{C_y} = 0,378$	$\frac{C_z}{C_y} = 0,621$	1
Число степеней свободы, ν	$l - 1 = 4$	$n - l = 95$	$n - 1 = 99$
Взвешенная дисперсия (девиата), σ^2	$\frac{C_x}{\nu_x} = 22,9$	$\frac{C_z}{\nu_z} = 1,58$	—
Критерий достоверности (вычисленный), F . . .	$\sigma_x^2 : \sigma_z^2 = 14,49$	—	—
Критерий достоверности F (табличное) при трех градациях p	$\begin{cases} 5,0 \\ 3,5 \\ 2,5 \end{cases}$	—	—

Этой сводной таблицей завершается дисперсионный анализ такого комплекса.

Следует иметь в виду, что при отсутствии таблиц Фишера со значением F критерий достоверности можно вычислить по формуле:

$$\theta = \frac{\nu_z - 2}{\nu_z} \cdot \frac{\sigma_x^2}{\sigma_z^2}.$$

Теоретическое значение θ равно 1. Если значение θ, взятое без единицы и отнесенное к значению его среднего квадратического отклонения σ_θ, будет больше или равно 3, то можно считать, что влияние фактора x на результативный признак y достоверно, т. е., если $\frac{\theta - 1}{\sigma_\theta} \geqslant 3$, то η_x^2 достоверно.

Значение σ вычисляется при этом по следующей формуле:

$$\sigma_\theta = \sqrt{\frac{2(\nu_z + \nu_x - 2)}{(\nu_z - 4)\,\nu_x}}.$$

Вычислим для приведенного примера значение достоверности указанным способом. Для этого подставляем в формулы θ и σ_θ нужные значения, т. е. ν_z, ν_x, σ_x^2, σ_z^2 и получаем:

$$\theta = \frac{\nu_z - 2}{\nu_z} \cdot \frac{\sigma_x^2}{\sigma_z^2} = \frac{95-2}{95} \cdot \frac{22,9}{1,58} = \frac{93}{95} \cdot 14,49 = 0,97 \cdot 14,49 = 14,05;$$

$$\sigma_\theta = \sqrt{\frac{2(95+4-2)}{(95-4)4}} = \sqrt{\frac{2 \cdot 97}{91 \cdot 4}} = \sqrt{\frac{194}{364}} = \sqrt{0,5328} = 0,72;$$

после этого вычисляем достоверность:

$$\frac{\theta - 1}{\sigma_\theta} = \frac{14,05-1}{0,72} = \frac{13,05}{0,72} = 18,1,$$

т. е. имеет место полная достоверность влияния фактора x на результативный признак y.

в. Обработка двухфакторного равномерного комплекса при малом числе наблюдений. Более сложными статистическими комплексами являются двухфакторные комплексы, в которых представлено влияние двух разных факторов, воздействующих на результативный признак. В многофакторных комплексах обработке подвергаются данные, относящиеся только к независимым друг от друга факторам; таковыми, например, могут быть температура и влажность воздуха, влияющие на урожайность, тип опыления и видовое различие пыльцы, использованные для изучения процессов опыления и т. п.

При обработке многофакторных, а следовательно, и двухфакторных комплексов техника обработки меняется в зависимости от того, имеется ли в комплексе равномерное, пропорциональное или неравномерное соотношение частот по факторам и их градациям; прежде чем приступать к обработке комплекса, это необходимо иметь в виду.

В двухфакторном комплексе на результативный признак оказывают влияние не только факторы A, B и остаточный фактор Z, но также и совместное действие обоих факторов, т. е. AB, величину которого можно вычислить в ходе дисперсионного анализа.

В связи с этим дисперсионный анализ должен дать следующие значения: C_y — общая дисперсия; C_x — общефакториальная дисперсия; C_A — дисперсия от воздействия фактора

A_1, C_B — дисперсия, вызываемая фактором B; C_{AB} — дисперсия, вызываемая совместным влиянием учтенных факторов A и B; C_z — остаточная дисперсия.

Если анализу подвергается двухфакторный комплекс с равномерным распределением частот и малым числом наблюдений, то дисперсии выражаются следующими формулами:

$$C_y = \Sigma V^2 - H, \quad \text{где} \quad H = \frac{(\Sigma V)^2}{n};$$

$$C_x = \Sigma h_x - H, \quad \text{где} \quad \Sigma h_x = \Sigma \frac{(\Sigma V_x)^2}{n_x};$$

$$C_A = \Sigma h_A - H, \quad \text{где} \quad \Sigma h_A = \Sigma \frac{(\Sigma V_A)^2}{n_A};$$

$$C_B = \Sigma h_B - H, \quad \text{где} \quad \Sigma h_B = \Sigma \frac{(\Sigma V_B)^2}{n_B};$$

$$C_{AB} = C_x - C_A - C_B;$$

$$C_z = \Sigma V^2 - \Sigma h_x.$$

Из перечисленных выше формул известно вычисление значений C_y, C_x, C_z (см. стр. 158); для вычисления же C_A, C_B, C_{AB} необходимо найти значение частных, т. е. факториальных значений Σh, а именно, величину Σh_A, полученную из обработки классов по признаку A, и величину Σh_B, получаемую аналогичным способом при обработке классов второго воздействующего признака B.

Дисперсия же совместного влияния обоих факторов C_{AB}, вычисляется обычным арифметическим вычитанием с подстановкой величин C_x, C_A, C_B, если комплексы имеют равномерный, или пропорциональный тип.

Разберем пример двухфакторного равномерного комплекса при малом числе наблюдений.

Пример. В опыте над мышами были группы самцов и самок, получавших рационы, различающиеся по питательности.

Требуется выявить долю влияния питательности рациона (фактор A) и пола (фактор B) мышей на плодовитость. Исходные данные представлены в таблице статистического комплекса (табл. 34).

Обрабатываем статистический комплекс по первичным данным для получения общей, факториальной и остаточной дисперсий известным уже способом, т. е. получаем строчку

	Самки, B_1			Самцы, B_2			Число градаций по $A(l_A = 3)$ по $B(l_B = 2)$
	Рацион нормальный по содержанию белка, A_1	Рацион, обогащенный белком, A_2	Рацион с минимальным содержанием белка, A_3	Рацион с нормальным содержанием белка, A_1	Рацион, обогащенный белком, A_2	Рацион с минимальным содержанием белка, A_3	
Плодовитость самок (число мышат в помете), V	6; 10	8; 11	5; 3	7; 8	9; 10	4; 5	$\Sigma V = 86$
V^2	36; 100	64; 121	25; 9	49; 64	81; 100	16; 25	$\Sigma V^2 = 690$
n_x	2	2	2	2	2	2	$n = 12$
ΣV_x	16	19	8	15	19	9	$\Sigma V_x = 86$
$(\Sigma V_x)^2$	256	361	64	225	361	81	—
$h_x = \frac{(\Sigma V_x)^2}{n_x}$	$\frac{256}{2} = 128$	$\frac{361}{2} = 180,5$	$\frac{64}{2} = 32$	$\frac{225}{2} = 112,5$	$\frac{361}{2} = 180,5$	$\frac{81}{2} = 40,5$	$\Sigma h_x = 674$
							$H = \frac{86^2}{12} = 7396 : 12 = 616,3$

комплекса при малом числе наблюдений

Обработка по факторам A и B	$n_{\text{фактор}}$	$\Sigma V_{\text{фактор}}$	$(\Sigma V_{\text{фактор}})^2$	$h_{\text{фактор}} = \frac{(\Sigma V_{\text{фактор}})^2}{n_{\text{фактор}}}$	$M_{\text{фактор}} = \frac{\Sigma V_{\text{фактор}}}{n_{\text{фактор}}}$
A_1	4	31	961	$\frac{961}{4}=240{,}2$	$M_{A_1}=\frac{31}{4}=7{,}75$
A_2	4	38	1444	$\frac{1444}{4}=361$	$M_{A_2}=\frac{38}{4}=9{,}5$
A_3	4	17	289	$\frac{289}{4}=72{,}2$	$M_{A_3}=\frac{17}{4}=4{,}25$
Показатели по фактору A	$n_A=12$	$\Sigma V_A=86$		$\Sigma h_A=673{,}4$	$M_A=\frac{86}{12}=7{,}16$
B_1	6	43	1849	$\frac{1849}{6}=308{,}1$	$M_{B_1}=\frac{43}{6}=7{,}17$
B_2	6	43	1849	$\frac{1849}{6}=308{,}1$	$M_{B_2}=\frac{43}{6}=7{,}17$
Показатели по фактору B	$n_B=12$	$\Sigma V_B=86$		$\Sigma h_B=616{,}2$	$M_B=\frac{86}{12}=7{,}17$

значений V; затем получаем строчку со значениями V^2; далее подсчитываем число наблюдений в каждой клетке и получаем строчку n_x; суммируем в каждой клетке значения вариантов и получаем строчку ΣV_x, возводим значение ΣV_x в каждой клетке в квадрат и получаем строчку $(\Sigma V_x)^2$, полученное значение $(\Sigma V_x)^2$ в каждой клетке делим на соответствующее этому значение n_x и получаем строчку величин h_x, суммирование которой дает нужное значение $\Sigma h_x = 674$.

Нужное значение H вычисляется так: $\frac{(\Sigma V_x)^2}{n} = \frac{86^2}{12}$. Оно оказалось равным 616,3.

Таким образом, находятся значения H и Σh_x, входящие в формулы общей, факториальной и остаточной дисперсий, т. е. C_y, C_x и C_z.

Для получения частных значений Σh_x, т. е. Σh_A, Σh_B и Σh_{AB}, соответствующих воздействующим факторам, производится обработка данных таблицы аналогичным способом. Эта обработка проделана в правой стороне табл. 34. Здесь для каждого фактора с учетом числа его классов (или градаций) осуществляются все необходимые расчеты для получения частных значений h. Одновременно эти расчеты позволяют вычислить частные средние арифметические по каждой градации обоих факторов ($M_{\text{части}}$ или $M_{\text{фактор}}$), что представлено данными в правом крайнем столбце таблицы.

По фактору A в таблице имеется три класса (A_1, A_2, A_3), по каждому из которых и получаются соответствующие строчки значений $n_A = 12$; $\Sigma V_A = 86$; $\Sigma h_A = 673{,}4$ и $M_A = 7{,}16$.

По фактору B имеется два класса: B_1 и B_2.

Аналогичная обработка по каждому из этих классов построчно дает значения $n_B = 12$; $\Sigma V_B = 86$; Σh_B и M_B, причем Σh_B оказалась равной 616,2.

Имея значения $\Sigma h_A = 673{,}4$ и $\Sigma h_B = 616{,}2$, можно вычислить частные дисперсии этих факторов, т. е. C_A и C_B, а по выражению $C_x - C_A - C_B$ определить последний элемент дисперсионного анализа — дисперсию, вызываемую совместным действием факторов C_{AB}.

Подставим в формулы дисперсии необходимые найденные значения:

$$C_y = \Sigma V^2 - H = 690 - 616{,}3 = 73{,}7;$$

$$C_x = \Sigma h_x - H = 674 - 616{,}3 = 57{,}7;$$

$$C_A = \Sigma h_A - H = 673{,}4 - 616{,}3 = 57{,}1;$$

$$C_B = \Sigma h_B - H = 616{,}2 - 616{,}3 \approx 0;$$

$$C_{AB} = C_x - C_A - C_B = 57{,}7 - 57{,}1 - 0 = 0{,}6;$$

$$C_z = \Sigma V^2 - \Sigma h_x = 690 - 674 = 16.$$

Общая дисперсия должна быть равна сумме частных дисперсий, в нашем примере:

$$C_y = C_A + C_B + C_{AB} + C_z = 57{,}1 + 0 + 0{,}6 + 16{,}0 = 73{,}7.$$

Первый этап дисперсионного анализа показал, что фактор B, т. е. пол животных, получивших тот или иной тип рациона, не влияет на изменчивость результативного признака (плодовитость), так как $C_B = 0$.

Продолжим дисперсионный анализ по общепринятой схеме, для чего вычислим все необходимые показатели, входящие в сводную таблицу дисперсионного анализа (табл. 35).

Таблица 35

Сводная таблица для двухфакторного комплекса (к примеру на стр. 170)

	A	B	AB	x	z	y
C	57,1	0	0,6	57,7	16,0	73,7
η^2	0,774	0	0,008	0,782	0,217	1,0
ν	2	1	2	5	6	11
σ^2	28,5	0	0,3	11,5	2,66	—
F (вычислен)	10,7	0	0,11	4,32	—	—
F (табличн.)	27,0; 10,9; 5,1	—	27,0; 10,9; 5,1	20,8; 8,8; 4,4	—	—

Находим долю влияния η^2 каждой частной дисперсии в общей дисперсии:

$$\eta^2_x = \frac{57{,}7}{73{,}7} = 0{,}782;$$

$$\eta^2_z = \frac{16{,}0}{73{,}7} = 0{,}217;$$

$$\eta^2_A = \frac{57{,}1}{73{,}7} = 0{,}774;$$

$$\eta^2_B = 0;$$

$$\eta^2_{AB} = \frac{0,6}{73,7} = 0,008;$$

$$\eta^2_y = 0,774 + 0,008 + 0,217 = 1,0.$$

По этим данным определяем связь между плодовитостью и типом рациона, вычисляя корреляционное отношение η:

$$\eta = \sqrt{0,774} = 0,88;$$

этот результат говорит о наличии большой связи.

Дальнейший анализ требует определить число степеней свободы для каждой дисперсии.

Для двухфакторных комплексов можно пользоваться следующими выражениями, дающими число степеней свободы:

$$\nu_y = n - 1; \qquad \nu_A = l_A - 1;$$
$$\nu_x = l_A \cdot l_B - 1; \quad \nu_B = l_B - 1;$$
$$\nu_z = n - l_A \cdot l_B; \quad \nu_{AB} = \nu_A \cdot \nu_B,$$

где n — число наблюдений по всему комплексу;

l_A — число классов по фактору A;

l_B — число классов по фактору B.

Исходя из этого, в указанном примере число степеней свободы будет для каждого комплекса следующим:

$$\nu_y = 12 - 1 = 11; \quad \nu_x = 3 \cdot 2 - 1 = 5; \quad \nu_z = 12 - 6 = 6;$$
$$\nu_A = 3 - 1 = 2; \quad \nu_B = 2 - 1 = 1; \quad \nu_{AB} = 2 \cdot 1 = 2.$$

Эти данные записываются в сводную таблицу дисперсионного анализа в строчку ν. После этого вычисляем:

$$\sigma^2 = \frac{C}{\nu}; \quad \sigma^2_x = \frac{57,7}{5} = 11,5; \quad \sigma^2_z = \frac{16}{6} = 2,66;$$

$$\sigma^2_A = \frac{57,1}{2} = 28,5; \quad \sigma^2_{AB} = \frac{0,6}{2} = 0,3.$$

Дальнейшие действия заключаются в отыскании показателя F для определения достоверности влияния факторов:

$$F_A = \frac{\sigma^2_A}{\sigma^2_z} = \frac{28,5}{2,66} = 10,7;$$

достоверное значение F по таблице Фишера при трех вероятностях соответственно равно: 27,0; 10,9; 5,1; следовательно, действие фактора A достоверно при $P=0,95$ и $P=0,99$.

$$F_{AB}=\frac{\sigma_{AB}^2}{\sigma_z^2}=\frac{0,3}{2,66}=0,11;$$

достоверное значение F по таблице равно: 27,0; 10,9; 5,1; следовательно, совместное влияние факторов A и B не достоверно;

$$F_x=\frac{\sigma_x^2}{\sigma_z^2}=\frac{11,5}{2,66}=4,32;$$

достоверное значение F по таблицам равно: 20,8; 8,8; 4,4; следовательно, дисперсия, возникающая от учтенных факторов, достоверна при низком уровне вероятности ($P=0,95$).

Если сопоставить эти выводы с полученными значениями частных средних арифметических M, представленных в последнем правом столбце таблицы комплекса, то оказывается, что сделанные выводы о достоверности совпадают с вычисленными значениями M.

Так, по фактору A (уровень кормления) плодовитость существенно различалась: при нормальном рационе (A_1) плодовитость M_{A_1} равнялась 7,75 голов мышат в помете; при рационе с повышенным содержанием белка (A_2) плодовитость M_{A_2} повысилась до 9,5 голов; при низком уровне питательности рациона (A_3) — снизилась до $M_{A_3}=4,25$ голов.

Разности между средними M_{A_1} и M_{A_2}; M_{A_2} и M_{A_3} можно считать достоверными, что подтвердилось полученными значениями достоверности влияния фактора A по показателям F_A (при $P=0,95$).

Если же сравнить частные средние M_B, полученные по этому фактору (пол родителей, получавших тот или иной тип рациона), то оказывается, что средняя плодовитость $M_{B_1}=M_{B_2}=7,17$ голов, т. е. она не различается по градациям этого признака и, следовательно, не зависит от того, какой пол родительского поколения был подвергнут воздействию разных рационов.

г. Обработка двухфакторного пропорционального комплекса при большом числе наблюдений. Дисперсионный анализ двухфакторного комплекса при большом числе наблюдений начинается с составления корреляционной решётки. Дальнейшая обработка сходна с той, как это делалось при однофакторном комплексе (см. стр. 161, 162).

Для получения дисперсий требуется вычислить в ходе обработки статистического комплекса значения:

$$S_2 = \Sigma p_y \cdot a_y^2; \quad H = \frac{S_1^2}{n} = \frac{(\Sigma p_y \cdot a_y)^2}{n}; \ \Sigma h_x; \ \Sigma h_A; \ \Sigma h_B.$$

Формулы дисперсий ничем принципиально не отличаются от формул двухфакторного комплекса для малой выборки, а техника обработки близка к той, которая проводилась при обработке однофакторного комплекса при большом числе наблюдений (см. стр. 161, 162).

Приведем формулы дисперсий двухфакторного комплекса при большой выборке:

$$C_y = S_2 - H, \quad \text{где} \quad S_2 = \Sigma p_y \cdot a_y^2 \quad \text{или} \quad S_2 = q_2 + q_1 + 2(q_4 + q_3),$$

$$\text{а величина } H = \frac{S_1^2}{n} = \frac{(\Sigma p_y a_y)^2}{n}, \quad \text{где} \quad S_1 = q_2 - q_1;$$

$$C_x = \Sigma h_x - H, \quad \text{где} \quad \Sigma h_x = \Sigma \left\{ \frac{(\Sigma p_{xy} a_y)^2}{p_x} \right\};$$

$$C_z = S_2 - \Sigma h_x;$$

$$C_A = \Sigma h_A - H, \quad \text{где} \quad \Sigma h_A = \Sigma \left\{ \frac{(\Sigma p_{xy} a_y)^2}{n_A} \right\};$$

$$C_B = \Sigma h_B - H, \quad \text{где} \quad \Sigma h_B = \Sigma \left\{ \frac{(\Sigma p_{xy} a_y)^2}{n_B} \right\};$$

$$C_{AB} = C_x - C_A - C_B.$$

Разберем ход дисперсионного анализа на следующем примере двухфакторного комплекса.

Пример. Выяснить влияние типа осеменения (A) овцематок романовской породы и типа кормления (B) баранов-производителей на многоплодие (y) овцематок этой породы. Осеменение было применено следующее: семенем одного производителя (A_1), смесью семени двух производителей (A_2). Кормление одной группы баранов-производителей осуществлялось по обычным рационам (B_1), а другой — по рационам с добавлением специальных препаратов, стимулирующих процесс сперматогенеза. В опыте по всем группам было 200 овцематок. Двухфакторный комплекс дал следующее пропорциональное распределение с соотношением частот 1 : 2 в A_1 и 1 : 2 в A_2 (табл. 36).

Таблица 36

Обработка двухфакторного пропорционального комплекса при большом числе наблюдений

Плодовитость (число голов), y	Факторы, x				n_y	a_y	Сумма накопленных частот	
	осеменение контрольное, A_1		осеменение смесью семени, A_2					
	контрольный рацион, B_1	подкормка препаратом, B_2	контрольный рацион, B_1	подкормка препаратом, B_2			$q_1=244$	$q_3=62$
1	12	15	15	20	62	-2	62	62
2	15	40	20	45	120	-1	182	—
3	2	2	3	6	13	0	—	—
4	1	—	1	4	6	$+1$	15	—
5	—	3	1	5	9	$+2$	9	9
n_x	30	60	40	80	210	—	$q_2=24$	$q_4=9$
$\Sigma p_{x\,y}\cdot a_y$	-38	-64	-47	-71	-220	$S_1=q_2-q_1=24-244=-220$		
$(\Sigma p_{x\,y}\cdot a_y)^2$	1444	4096	2209	5041	—	$S_2=q_2+q_1+2(q_4+q_3)=$ $=24+224+2(9+62)=390$		
$h_x=\dfrac{(\Sigma p_{x\,y}\cdot a_y)^2}{n_x}$	$\frac{1444}{30}=$ $=48,13$	$\frac{4096}{60}=$ $=68,27$	$\frac{2209}{40}=$ $=55,22$	$\frac{5041}{80}=$ $63=,01$	$\Sigma h_x=234,64$			

Для получения ряда $\Sigma p_{xy} \cdot a_y$ производим умножение по столбцам каждого значения частот (p_{xy}) в клетках таблицы на соответствующее условное отклонение a_y:

1-й столбец: $12 \cdot (-2) = -24$; $15 \cdot (-1) = -15$; $2 \cdot 0 = 0$; $1 \cdot 1 = 1$; $\Sigma p_{xy} a_y = -38$.

2-й столбец: $15 \cdot (-2) = -30$; $40 \cdot (-1) = -40$: $2 \cdot 0 = 0$; $3 \cdot 2 = 6$; $\Sigma p_{xy} a_y = -64$.

3-й столбец: $15 \cdot (-2) = -30$; $20 \cdot (-1) = -20$; $3 \cdot 0 = 0$; $1 \cdot 1 = 1$; $1 \cdot 2 = 2$; $\Sigma p_{xy} a_y = -47$.

4-й столбец: $20 \cdot (-2) = -40$; $45 \cdot (-1) = -45$; $6 \cdot 0 = 0$; $4 \cdot 1 = 4$; $5 \cdot 2 = 10$; $\Sigma p_{xy} a_y = -71$.

Общая сумма $\Sigma(\Sigma p_{xy} \cdot a_y) = -220$ дает значение S_1, при помощи которого определяется величина H:

$$H = \frac{S_1^2}{n} = \frac{(-220)^2}{210} = \frac{48\,400}{210} = 230{,}5.$$

Следующая строчка таблицы получается путем возведения в квадрат значения $\Sigma p_{xy} \cdot a_y$ по каждому столбцу.

Деление значений $(\Sigma p_{xy} \cdot a_y)^2$ по каждому столбцу на соответствующее значение n_x дает величины h_x, от суммирования которых по строчке получается величина Σh_x, равная 234,64.

Проведенная обработка позволяет получить дисперсии C_y, C_x, C_z.
Найдем их величины:

$$C_y = S_2 - H = 390 - 230{,}5 = 159{,}5;$$

$$C_x = \Sigma h_x - H = 234{,}64 - 230{,}5 = 4{,}14;$$

$$C_z = S_2 - \Sigma h_x = 390 - 234{,}64 = 155{,}36.$$

Для получения дисперсий по каждому из факторов и их совместному действию, т. е. C_A, C_B и C_{AB}, требуется произвести аналогичную обработку по каждому из факторов. Для этого пользуются следующей подсобной таблицей, составляющей как бы продолжение таблицы статистического комплекса:

по фактору А

Градации по A	n_A	$\Sigma p_{x \cdot y} \cdot a_y$	$(\Sigma p_{x \cdot y} \cdot a_y)^2$	$h_A = \frac{(\Sigma p_{x \cdot y} \cdot a_y)^2}{n_A}$
A_1	90	$+\begin{array}{r}-38\\-64\\\hline-102\end{array}$	$(-102)^2 = 10404$	$\frac{10404}{90} = 115{,}6$
A_2	120	$+\begin{array}{r}-47\\-71\\\hline-118\end{array}$	$(-118)^2 = 13924$	$\frac{13924}{120} = 116{,}0$
	210	-220	—	$\Sigma h_A = 231{,}6$

Отсюда:

$$C_A = \Sigma h_A - H = 231.6 - 230{,}5 = 1{,}1.$$

по фактору B

Градации по B	n_B	$\Sigma p_{x \cdot y}\, a_y$	$(\Sigma p_{x\ y} \cdot a_y)^2$	$h_B = \frac{(\Sigma p_{x \cdot y} \cdot a_y)^2}{n_B}$
B_1	70	$+\begin{array}{r}-38\\-47\\\hline-85\end{array}$	$(-85)^2=7225$	$\frac{7225}{70} = 103{,}2$
B_2	140	$+\begin{array}{r}-64\\-71\\\hline-135\end{array}$	$(-135)^2=18225$	$\frac{18225}{140} = 130{,}2$
	210	-220	—	$\Sigma h_B = 233{,}4$

Отсюда:

$$C_B = \Sigma h_B - H = 233{,}4 - 230{,}5 = 2{,}9.$$

Находим дисперсию совместного действия факторов:

$$C_{AB} = C_x - C_A - C_B = 4{,}14 - 1{,}1 - 2{,}9 = 0{,}14.$$

Завершаем анализ составлением сводной таблицы дисперсионного анализа (табл. 37).

Здесь n — общее число наблюдений в таблице;

l_A — число классов по фактору A;

l_B — число классов по фактору B.

Табличные значения F_A : 11,2; 6,8; 3,9; ($\nu_A = 1$, $\nu_z = 206$;)

F_B : 11,2; 6,8; 3,9; ($\nu_B = 1$, $\nu_z = 206$):

F_{AB} : 11,2; 6,8; 3,9 ($\nu_{AB} = 1$, $\nu_z = 206$);

F_x : 5,6; 3,9; 2,4; ($\nu_x = 3$, $\nu^z = 206$):

Из полученных данных видно, что влияние изученных факторов х, A и B на плодовитость очень мало, (2,59; 0,63 и 1,81%), а достоверное же влияние оказалось на уровне $P=0{,}95$ для фактора B. Следовательно, на плодовитость оказывает влияние тип кормления самцов.

д. Обработка трехфакторного статистического равномерного комплекса ничем существенным не отличается от того, как это делается для двухфакторных комплексов. Составление таблицы производится тем же путем, когда результативный признак выписывается по горизонтали, т. е. по строчкам, а факториальные показатели — по вертикали, т. е. по столбцам.

Кроме C_y, C_x, C_z, C_A, C_B и C_{AB} вычисляются дисперсии, обусловленные наличием воздействия третьего фактора C, т. е. дисперсии C_C, C_{AC}, C_{BC} и C_{ABC}.

Формулы трехфакторной дисперсии при малом числе наблюдений существуют следующие:

$$C_y = \Sigma V^2 - H, \quad \text{где} \quad H = \frac{(\Sigma V)^2}{n};$$

$$C_x = \Sigma h_x - H, \quad \text{где} \quad \Sigma h_x = \Sigma \frac{(\Sigma V)^2}{n_x};$$

$$C_z = \Sigma V^2 - \Sigma h_x;$$

$$C_A = \Sigma h_A - H, \quad \text{где} \quad \Sigma h_A = \Sigma \frac{(\Sigma V)^2}{n_A};$$

$$C_B = \Sigma h_B - H, \quad \text{где} \quad \Sigma h_B = \Sigma \frac{(\Sigma V)^2}{n_B};$$

$$C_C = \Sigma h_C - H, \quad \text{где} \quad \Sigma h_C = \Sigma \frac{(\Sigma V)^2}{n_C};$$

Таблица 37

Св дная таблица дисперсионного анализа двухфакторного комплекса (к примеру на стр. 177)

	A	B	AB	x	z	y
C	1,1	2,9	0,14	4,14	155,36	159,5
$\eta_i^2=\frac{C_{фактор}}{C_y}$	$\frac{1,1}{159,5}=$ $=0,0063=$ $=0,63\%$	$\frac{2,9}{159,5}=$ $=0,0181=$ $=1,81\%$	$\frac{0,14}{159,5}=$ $=0,00015=$ $=0,001\%$	$\frac{4,14}{159,5}=$ $=0,0259=$ $=2,59\%$	$\frac{155,36}{159,5}=$ $=0,974=$ $=97\ 4\%$	$\frac{159,5}{1\ 9,5}=$ $=1,0=$ $=10\ \%$
ν	$l_A-1=2-1=1$	$l_B-1=2-1=1$	$\nu_A\cdot\nu_B=1\cdot 1=1$	$l_A\cdot l_B-1=$ $=2\cdot 2-1=3$	$n-l_A$ $=210-2\cdot\ =206$	$n-1=$ $210-1=20$
$\sigma^2=\frac{C}{\nu}$	$\frac{1,1}{1}=1,1$	$\frac{2,9}{1}=2,9$	$\frac{0,14}{1}=0,14$	$\frac{4,14}{3}=1,38$	$\frac{155,36}{206}=0,754$	—
$F=\frac{\sigma^2_{фактор}}{\sigma^2_z}$	$\frac{1,1}{0,754}=1,32$	$\frac{2,9}{0,754}=3,84$	$\frac{0,14}{0,754}=0,31$	$\frac{1,38}{0,754}=1,83$	—	—

$$C_{AB} = \Sigma h_{AB} - H - C_A - C_B, \quad \text{где} \quad \Sigma h_{AB} = \Sigma \frac{(\Sigma V)^2}{n_{AB}};$$

$$C_{AC} = \Sigma h_{AC} - H - C_A - C_C, \quad \text{где} \quad \Sigma h_{AC} = \Sigma \frac{(\Sigma V)^2}{n_{AC}},$$

$$C_{BC} = \Sigma h_{BC} - H - C_B - C_C, \quad \text{где} \quad \Sigma h_{BC} = \Sigma \frac{(\Sigma V)^2}{n_{BC}};$$

$$C_{ABC} = C_x - C_A - C_B - C_C - C_{AB} - C_{AC} - C_{BC}.$$

Число степеней свободы для каждой дисперсии вычисляется по формулам, из которых часть уже были использованы при обработке двухфакторного комплекса:

$\nu_v = n - 1$; $\nu_x = l_A l_B l_C - 1$ (где l_A, l_B и l_C — число классов по каждому фактору);

$\nu_z = n - l_A l_B l_C$; $\quad \nu_A = l_A - 1$; $\quad \nu_B = l_B - 1$; $\quad \nu_C = l_C - 1$;

$\nu_{AB} = \nu_A \nu_B$; $\quad \nu_{AC} = \nu_A \nu_C$; $\quad \nu_{BC} = \nu_B \nu_C$;

$\nu_{ABC} = \nu_A \nu_B \nu_C$.

Если трехфакторный комплекс с равномерным, или пропорциональным соотношением частот имеет большое число наблюдений, то составляется корреляционная решетка, которая обрабатывается по тому же принципу, как это было показано в отношении двухфакторного комплекса, а формулы дисперсий можно использовать те, которые приведены в данной главе.

Б. Неравномерные статистические комплексы

Обработка неравномерных комплексов, т. е. таких, у которых частоты по градациям одного фактора не равны по градациям другого фактора и не пропорциональны, будет отличаться от обработки равномерных и пропорциональных (ортогональных) комплексов.

В неравномерных комплексах $C_A + C_B + C_{AB} \neq C_X$, т. е. дисперсия по организованным изучаемым факторам (C_X) не равна сумме частных дисперсий (C_A и C_B) и дисперсии от их взаимного действия (C_{AB}), также и $\eta_A^2 + \eta_B^2 + \eta_{AB}^2 \neq \eta_X^2$, в то время, как в равномерных комплексах всегда $C_A + C_B + C_{AB} = C_X$ и $\eta_A^2 + \eta_B^2 + \eta_{AB}^2 = \eta_X^2$.

Для неравномерных комплексов остается в силе равенство: $C_y = C_x + C_z$, имевшее место у равномерных комплексов.

Таким образом, в неравномерных комплексах изменение системы обработки материала связано с вычислением частных дисперсий (например, в трехфакторном комплексе с вычислением дисперсий *А, В, С, АВ, ВС, АС, АВС*), в то время, как общефакториальная дисперсия C_x, остаточная дисперсия C_z и общая дисперсия C_y вычисляются теми же способами, как это делалось в равномерных и пропорциональных комплексах. Неравенство $C_A + C_B + C_{AB} \neq C_x$ свидетельствует о том, что в неравномерных комплексах возникает еще дополнительная дисперсия за счет неравномерного распределения частот по градациям факторов.

Для устранения этого пользуются различными приёмами, детально разобранными в книге Н. А. Плохинского[14]. Остановимся только на одном приёме, предложенном Ю. Л. Поморским, когда неравномерность комплекса устраняется путем усреднения частот по градациям (или классам) факторов. При этом получают по каждой градации каждого фактора частное среднее арифметическое из значений вариантов (т. е. $M_{\text{частн}}$), затем возводят полученные частные средние в квадрат и подставляют эти значения в формулы, несколько отличающиеся от тех формул, которыми ранее пользовались для вычисления C_A, C_B, C_{AB}.

Формулы общефакториальной и частных дисперсий для этого приёма применяются следующие:

$$C_X = n\left(\frac{\Sigma M_x^2}{l_A l_B} - M^2_{\text{общая}}\right),$$

где n — число наблюдений в комплексе;

ΣM_x^2 — сумма квадратов элементарных средних арифметических по всем градациям суммарного фактора x;

$M^2_{\text{общая}}$ — квадрат средней арифметической, вычисленной для всего комплекса;

l_A — число градаций по фактору A;

l_B — число градаций по фактору B;

частные дисперсии в неравномерных комплексах имеют такие выражения.

$$C_A = n\left(\frac{\Sigma M_A^2}{l_A} - M^2_{\text{общая}}\right).$$

---

[14] Н. А. Плохинский. Дисперсионный анализ. Новосибирск, Изд-во Сиб. отделения АН СССР, 1960.

Таблица 38

Обработка двухфакторного неравномерного комплекса при малом числе наблюдений

	Концентратное кормление, A_1		Полуконцентратное кормление, A_2		Число градаций по факторам: $l_A=2$, $l_B=2$
	свиноматки беркширы, B_1	свиноматки крупные белые, B_2	свиноматки беркширы, B_1	свиноматки крупные белые, B_2	суммарные данные по строкам
Число желтых тел, V	15; 12; 10; 16	20; 18; 17	18; 15; 16	22; 20; 17; 15; 20	
ΣV^2	$15^2+12^2+10^2+16^2=725$	$20^2+18^2+17^2=1013$	$18^2+15^2+16^2=805$	$22^2+20^2+17^2+15^2+20^2=1798$	$\Sigma(\Sigma V)^2=4341$
n_x	4	3	3	5	$n=15$
ΣV	53	55	49	94	$\Sigma V=231$
$(\Sigma V)^2$	$53^2=2809$	$55^2=3025$	$49^2=2401$	$94^2=8836$	—
$h_x=\frac{(\Sigma V)^2}{n_x}$	$\frac{2809}{4}=702,2$	$\frac{3025}{3}=1008,3$	$\frac{2401}{3}=800,3$	$\frac{8836}{5}=1767,2$	$\Sigma h_x=4278$
$M_x=\frac{\Sigma V}{n_x}$	$\frac{53}{4}=13,25$	$\frac{55}{3}=18,33$	$\frac{49}{3}=16,33$	$\frac{94}{5}=18,80$	$\Sigma M_x=66,71$
M_x^2	$13,25^2=175,56$	$18,33^2=335,98$	$16,33^2=266,67$	$18,80^2=353,44$	$\Sigma M_x^2=1131,65$

где ΣM_A^2 — сумма квадратов средних арифметических по градациям фактора A;

$$C_B = n\left(\frac{\Sigma M_B^2}{l_B} - M_{общая}^2\right),$$

где ΣM_B^2 — сумма квадратов средних арифметических по градациям фактора B;

$$C_{AB} = C_X - C_A - C_B.$$

Последним равенством уже можно пользоваться, так как произведено усреднение частот по градациям путем получения $M_{частн}^2$; C_y и C_z вычисляются по обычным, уже известным, формулам, а именно:

$$C_z = \Sigma V^2 - \Sigma h_x \qquad \text{и} \qquad C_y = C_x + C_z.$$

а. Обработка двухфакторного неравномерного комплекса разбирается на следующем примере.

Пример. Выяснить влияние типа кормления (фактор A) и породы (фактор B) на потенциальную плодовитость свиноматок, определяемую числом желтых тел на яичниках. В опыте были свиноматки породы беркшир и крупной белой. Типы кормления, использованные в опыте на матках обеих пород, — концентратный и полуконцентратный.

Итоговые данные опыта сведены в таблице статистического комплекса (табл. 38).

Дальнейшая аналогичная обработка осуществляется по каждому частному фактору A и B; расчёты этой обработки указаны в подсобной табл. 39.

Вычислим общую среднюю арифметическую, пользуясь суммой элементарных средних ΣM_x

$$M_{общая} = \frac{\Sigma M_x}{l_A l_B} = \frac{66,71}{2 \cdot 2} = \frac{66,71}{4} = 16,68;$$

возведем ее в квадрат:

$$M_{общая}^2 = 16,68^2 = 278,22.$$

Вычисляем все необходимые дисперсии C:

$$C_x = n\left(\frac{\Sigma M_x^2}{l_A l_B} - M_{общая}^2\right) = 15\left(\frac{1131,65}{2 \cdot 2} - 278,22\right) =$$

$$= 15 \cdot (282,91 - 278,22) = 15 \cdot 4,69 = 70,35;$$

Обработка данных по частным факторам (к примеру на стр. 184)

Фактор	A_1	A_2		B_1	B_2	Σ
Число средних арифметических, входящих в градации, l	2	2	4	2	2	**4**
ΣM_X	31,58	35,13	66,71	29,58	37,13	66,71
$M_{частн} = \frac{\Sigma M_X}{l}$	$\frac{31,58}{2} = 15,79$	$\frac{35,13}{2} = 17,56$	—	$\frac{29,58}{2} = 14,79$	$\frac{37,13}{2} = 18,56$	—
$M^2_{частн}$	$15,79^2 = 249,32$	$17,56^2 = 308,35$	$\Sigma M_A^2 = 557,67$	$14,79^2 = 218,74$	$18,56^2 = 344,47$	$\Sigma M_B^2 = 563,21$

$$C_A = n\left(\frac{\Sigma M_A^2}{l_A} - M^2_{общая}\right) = 15\left(\frac{557,67}{2} - 278,22\right) =$$

$$= 15\cdot(278,83 - 278,22) = 15\cdot 0,61 = 9,15;$$

$$C_B = n\left(\frac{\Sigma M_B^2}{l_B} - M^2_{общая}\right) = 15\left(\frac{563,21}{2} - 278,22\right) =$$

$$= 15\,(281,60 - 278,22) = 15\cdot 3,38 = 50,70,$$

$$C_{AB} = C_X - C_A - C_B = 70,35 - 9,15 - 50,70 =$$

$$= 70,35 - 59,85 = 10,50;$$

$$C_Z = \Sigma V^2 - \Sigma h_X = 4341 - 4278 = 63;$$

$$C_y = C_X + C_Z = 70,35 + 63 = 133,35.$$

Последующие расчёты осуществляются в соответствии с графами сводной таблицы дисперсионного анализа обычным способом (табл. 40).

Сводная таблица дисперсионного анализа (к примеру на стр. 184)

	A	B	AB	X	Z	Y
C	9,15	50,70	10,50	70,35	63,0	133,35
$\eta^2 = \frac{C_{факт}}{C_y}$	0,0686	0,3802	0,07874	0,5275	0,4725	1,0
ν	$l_A - 1 = 2 - 1 = 1$	$l_B - 1 = 2 - 1 = 1$	$\nu_A \nu_B = 1 \cdot 1 = 1$	$l_A \cdot l_B = 2 \cdot 2 = 4$	$n - l_A \cdot l_B = 15 - 2 \cdot 2 = 11$	$n - 1 = 15 - 1 = 14$
$\sigma^2 = \frac{C}{\nu}$	9,15	50,70	10,50	17,59	5,727	—
$F = \frac{\sigma^2_{факт}}{\sigma^2_Z}$	$\frac{9,15}{5,727} = 1,59$	$\frac{50,70}{5,727} = 8,8$	$\frac{10,50}{5,727} = 1,8$	$\frac{17,59}{5,727} = 3,07$	—	—
Табличное значение, F	19,7 9,7 4,8	19,7 9,7 4,8	19,7 9,7 4,8	10,4 5,7 3,4	—	—

Из сводной таблицы обработанного неравномерного комплекса видно, что доля суммарного влияния изученных факторов η^2_x (порода и тип рациона) на первичную плодовитость составляет 52,75% от общей дисперсии, но достоверность этого отсутствует, так как вычисленное значение F меньше табличного значения F по всем градациям. Влияние фактора A (тип рациона) выражается в 6,86%, но оно также недостоверно. Влияние фактора B (порода свиноматок) выражается 38,02% от общей дисперсии, что достоверно при $P = 0,95$.

б. Обработка трехфакторного неравномерного комлекса. Удобную схему дисперсионного анализа при трехфакторном неравномерном комплексе дает Н. А. Плохинский в книге

«Дисперсионный анализ», которую автор предлагает взять в качестве стандартного решения такого рода комплексов.

Для трехфакторного комплекса составляется обычным порядком таблица статистического комплекса, в которую построчно вносятся следующие данные:

$$V,\ n_x,\ \Sigma V,\ (\Sigma V)^2,\ h_x = \frac{(\Sigma V)^2}{n_x},\ \Sigma V^2,\ M_x,\ M^2_x.$$

Затем в подсобных таблицах проводятся расчёты для каждого частного фактора, а также и их сочетаний, т. е. A, B, C, AB, AC, BC. При этом вычисляются такие показатели, как ΣM_x, $M_{частн}$ и $M^2_{частн}$.

Полученные построчно и по столбцам суммарные значения этих показателей используются в следующих формулах:

$$H = \frac{(\Sigma V)^2}{n}; \quad M_{общая} = \frac{\Sigma M_X}{l_A l_B l_C}; \quad M^2_{общая} = \frac{(\Sigma M_X)^2}{(l_A l_B l_C)^2};$$

$$h_X = \frac{\Sigma M^2_X}{l_A l_B l_C}; \quad h_A = \frac{\Sigma M^2_A}{l_A}; \quad h_B = \frac{\Sigma M^2_B}{l_B}; \quad h_C = \frac{\Sigma M^2_C}{l_C};$$

$$h_{AB} = \frac{\Sigma M^2_{AB}}{l_A l_B}; \quad h_{AC} = \frac{\Sigma M^2_{AC}}{l_A l_C}; \quad h_{BC} = \frac{\Sigma M^2_{BC}}{l_B l_C}.$$

Как уже отмечалось, вследствие неравномерного распределения частот в неравномерных комплексах создается дополнительная дисперсия, приводящая к неравенству между общефакториальной дисперсией C_x и суммой частных дисперсий. Для трехфакторного комплекса это неравенство выражается следующим образом:

$$C_X \neq C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC}.$$

Общая же дисперсия C_y, как и в равномерных комплексах, *получается из равенства:* $C_y = C_x + C_z$.

Общефакториальная дисперсия C_x, вычисленная в неравномерном комплексе обычным путем по формуле: $C_x = \Sigma h - H$, являясь истинной дисперсией, оказывается несколько меньше, чем та общефакториальная дисперсия, которая в неравномерных комплексах получается суммированием частных дисперсий. Обозначим общефакториальную и частные дисперсии, получаемые в неравномерных комплексах, через C'. Следовательно,

$$C_X < C'_X;$$

$$C'_X = \Sigma C'_{частн} = C'_A + C'_B + C'_C + C'_{AB} + C'_{AC} + C'_{BC} + C'_{ABC}.$$

В этом равенстве все частные дисперсии также отличаются от соответствующих им дисперсий, которые были бы получены при равномерном *распределении частот*.

Поэтому при обработке неравномерного комплекса *требуется ввести на* частные дисперсии такую поправку (κ), которая корректировала бы эти частные дисперсии на наличие в комплексе неравномерности в распределении частот. Для этого значения частных дисперсий должны быть умножены на поправочный коэффициент κ. Таким поправочным коэффициентом κ может служить частное, получаемое от деления C_x на C'_x, т. е. $\kappa = \frac{C_x}{C'_x}$. Следовательно, для указанной корректировки частных дисперсий требуется осуществить такое *вычисление:*

$$C_A = C'_A\kappa; \quad C_B = C'_B\kappa; \quad C_C = C'_C\kappa;$$

$$C_{AB} = C'_{AB}\kappa; \quad C_{AC} = C'_{AC}\kappa;$$

$$C_{BC} = C'_{BC}\kappa; \quad C_{ABC} = C'_{ABC}\kappa.$$

В результате этого получаются окончательные величины истинного значения частных дисперсий. Общефакториальная и остаточная дисперсии вычисляются общепринятым способом, а именно:

$$C_X = \Sigma h_X - H \quad \text{и} \quad C_Z = \Sigma V^2 - \Sigma h_X.$$

Имея в виду эти общие замечания *о ходе обработки трехфакторного неравномерного комплекса, разберем последовательность этих расчётов на конкретном примере.*

Пример. Провести дисперсионный анализ на выявление влияния различных доз облучения мышей различных линий, подвергавшихся облучению в разном возрасте. Исходные данные этого опыта представлены в табл. 41.

Вычислим значения H, C_x, C_z, C_y и $M^2_{общая}$:

$$H = \frac{(\Sigma V)^2}{n} = \frac{146^2}{40} = \frac{21316}{40} = 532{,}9;$$

$$C_X = \Sigma h_X - H = 593{,}8 - 532{,}9 = 60{,}9;$$

$$C_Z = \Sigma V^2 - \Sigma h_X = 680 - 593{,}8 = 86{,}2;$$

$$C_y = \Sigma V^2 - H = 680 - 532{,}9 = 147{,}1.$$

Находим для всего комплекса общую среднюю арифметическую $M_{общая}$ и ее квадрат $M^2_{общая}$:

Таблица 41

Обработка трехфакторного неравномерного комплекса при малом числе наблюдений

	Резистентная линия мышей, A_1				Нерезистентная линия мышей, A_2				Число градаций по факторам равно: $l_A=2$, $l_B=2$, $l_C=2$,
	облученные максимальной дозой, B_1		облученные минимальной дозой, B_2		облученные максимальной дозой, B_1		облученные минимальной дозой, B_2		
	до половой зрелости, C_1	в половой зрелости, C_2	до половой зрелости, C_1	в половой зрелости, C_2	до половой зрелости, C_1	в половой зрелости, C_2	до половой зрелости, C_1	в половой зрелости, C_2	суммарные данные по строкам
Число мышат в помете, V	2 5 3 1 4	4 6 3 1 5	4 3 6	6 5 7 8	0 1 1 3 4 2	3 3 2 1 5	1 4 4 6 3 3 7	5 3 6 3 3	
n_x	5	5	3	4	6	5	7	5	$\Sigma n_x=40$
ΣV	15	19	13	26	11	14	28	20	$\Sigma V=146$
$(\Sigma V)^2$	225	361	169	676	121	196	784	400	—
$h_x=\frac{(\Sigma V)^2}{n_x}$	$\frac{225}{5}=45$	$\frac{361}{5}=72,2$	$\frac{169}{3}=56,3$	$\frac{676}{4}=169$	$\frac{121}{6}=20,1$	$\frac{196}{5}=39,2$	$\frac{784}{7}=112$	$\frac{400}{5}=80$	$\Sigma h_x=593,8$
ΣV^2	55	87	61	174	31	48	136	88	$\Sigma\Sigma V^2=680$
$M_x=\frac{\Sigma V}{n_x}$	$\frac{15}{5}=3$	$\frac{19}{5}=3,8$	$\frac{13}{3}=4,33$	$\frac{26}{4}=6,5$	$\frac{11}{6}=1,8$	$\frac{14}{5}=2,8$	$\frac{28}{7}=4$	$\frac{20}{5}=4$	$\Sigma M_x=30,23$
M_X^2	9	14,44	18,75	42,25	3,24	7,84	16,0	16,0	$\Sigma M_X^2=127,52$

$$M_{общая} = \frac{\Sigma M_X}{l_A l_B l_C} = \frac{30,23}{2\ 2\ 2} = \frac{30,23}{8} = 3,78,$$

откуда:

$$M^2_{общая} = 3,78^2 = 14,288.$$

После нахождения общей, общефакториальной и остаточной дисперсий переходим к вычислению частных дисперсий и их сочетаний. Обработку проводим с использованием подсобной таблицы по каждому из факторов и по каждому сочетанию (табл. 42).

Подставляем полученные значения $M^2_{общая}$ и $M^2_{частн}$ в соответствующие формулы h и частных дисперсий; при этом получаем увеличенные значения частных дисперсий $C_{частн}$, не исправленные на неравномерность комплекса:

$$h_X = \frac{\Sigma M^2_X}{l_A l_B l_C} = \frac{127,52}{2\ 2\ 2} = 15,94;$$

$$h_A = \frac{\Sigma M^2_A}{l_A} = \frac{29,344}{2} = 14,672;$$

$$C'_A = n(h_A - M^2_{общая}) = 40(14,672 - 14,288) = 40 \cdot 0,384 = 15,36;$$

$$h_B = \frac{\Sigma M^2_B}{l_B} = \frac{30,3066}{2} = 15,1533;$$

$$C'_B = n(h_B - M^2_{общая}) = 40(15,1533 - 14,288) = 40 \cdot 0,865 = 34,60$$

$$h_C = \frac{\Sigma M^2_C}{l_C} = \frac{29,047}{2} = 14,523;$$

$$C'_C = n(h_C - M^2_{общая}) = 40 \cdot (14,523 - 14,288) = 40 \cdot 0,235 = 9,4;$$

$$h_{AB} = \frac{\Sigma M^2_{AB}}{l_A l_B} = \frac{62,12}{2\ 2} = 15,53;$$

$$C'_{AB} = n(h_{AB} - h_A - h_B + M^2_{общая}) =$$
$$= 40 \cdot (15,53 - 14,672 - 15,139 + 14,288) = 40 \cdot 0,007 = 0,28;$$

$$h_{AC} = \frac{\Sigma M^2_{AC}}{l_A l_C} = \frac{59,88}{2.2} = 14,97;$$

Таблица 42

Обработка данных по частным факторам (к примеру на стр. 190)

Факторы	Число средних арифметических в градации	ΣM_x по градациям	$M_{частн} = \frac{\Sigma M_x}{i}$	$M^2_{частн} = \left(\frac{\Sigma M_x}{i}\right)^2$
A_1	4	3+3,8+4,33+6,5= =17,63	17,63:4=4,407	$4,407^2=19,422$
A_2	4	1,8+2,8+4+4=12,60	12,60:4=3,150	$3,15^2=9,922$
				$\Sigma M^2_A=29,344$
B_1	4	3+3,8+1,8+2,8= =11,40	11,40:4=2,85	$2,85^2=8,1225$
B_2	4	4,33+6,5+4+4= =18,83	18,83:4=4,71	$4,71^2=22,1841$
				$\Sigma M^2_B=30,3066$
C_1	4	3+4,33+1,8+4= =13,13	13,13:4=3,282	$3,282^2=10,771$
C_2	4	3,8+6,5+2,8+4= =17,10	17,10:4=4,275	$4,275^2=18,276$
				$\Sigma M^2_C=29,047$
A_1B_1	2	3+3,8=6,80	6,80:2=3,40	$3,40^2=11,56$
A_1B_2	2	4,33+6,5=10,83	10,83:2=5,41	$5,41^2=29,27$
A_2B_1	2	1,8+2,8=4,60	4,60:2=2,30	$2,30^2=5,29$
A_2B_2	2	4,0+4,0=8,00	8,00:2=4,00	$4,00^2=16,00$
				$\Sigma M^2_{AB}=62,12$
A_1C_1	2	3+4,33=7,33	7,33:2=3,66	$3,66^2=13,39$
A_1C_2	2	3,8+6,5=10,30	10,30:2=5,15	$5,15^2=26,52$
A_2C_1	2	1,8+4,0=5,80	5,80:2=2,90	$2,90^2=8,41$
A_2C_2	2	2,8+4,0=6,80	6,80:2=3,40	$3,40^2=11,56$
				$\Sigma M^2_{AC}=59,88$
B_1C_1	2	3+1,8=4,80	4,80:2=2,40	$2,40^2=5,76$
B_1C_2	2	3,8+2,8=6,60	6,60:2=3,30	$3,30^2=10,89$
B_2C_1	2	4,33+4,0=8,33	8,33:2=4,16	$4,16^2=17,31$
B_2C_2	2	6,50+4,0=10,50	10,50:2=5,25	$5,25^2=27,56$
				$\Sigma M^2_{BC}=61,52$

$$C'_{AC} = n\,(h_{AC} - h_A - h_C + M^2_{общая}) =$$

$$= 40 \cdot (14{,}97 - 14{,}672 - 14{,}523 + 14{,}288) =$$

$$= 40 \cdot 0{,}063 = 2{,}52;$$

$$h_{BC} = \frac{\Sigma M^2_{BC}}{l_B\, l_C} = \frac{61{,}52}{2.2} = 15{,}38;$$

$$C'_{BC} = n\,(h_{BC} - h_B - h_C + M^2_{общая}) =$$

$$= 40 \cdot (15{,}38 - 15{,}139 - 14{,}523 + 14{,}288) = 40 \cdot 0{,}006 = 0{,}24.$$

По полученным значениям h вычисляем дисперсию сочетания трех факторов:

$$C'_{ABC} = n\,(h_X + h_A + h_B + h_C - h_{AB} - h_{AC} - h_{BC} - M^2_{общая}) =$$

$$= 40 \cdot (15{,}94 + 14{,}672 + 15{,}1533 + 14{,}523 - 15{,}53 - 14{,}97 -$$

$$- 15{,}38 - 14{,}288) = 40 \cdot 0{,}12 = 4{,}8.$$

По значениям частных C' вычисляются значения C'_x:

$$C'_X = \Sigma C'_{частн} = C'_A + C'_B + C'_C + C'_{AB} + C'_{AC} + C'_{BC} + C'_{ABC} =$$

$$= 15{,}36 + 34{,}60 + 9{,}40 + 0{,}28 + 2{,}52 + 0{,}24 + 4{,}8 = 67{,}20.$$

Таким образом, в результате проведенной обработки комплекса получены два значения общефакториальной дисперсии:

$$C'_X = 67{,}20 \text{ и } C_X = 60{,}90.$$

Разница в значениях дисперсий есть результат неравномерного распределения частот. Для устранения влияния этого распределения на дисперсии вычисляем поправочный коэффициент k:

$$k = \frac{C_X}{C'_X} = \frac{60{,}90}{67{,}20} = 0{,}906.$$

На этот коэффициент следует умножить все частные дисперсии $C'_{частн}$.

В результате введения такой поправки получаются значения дисперсий, которые были бы получены в условиях равномерного распределения частот в комплексе, а именно, корректированные значения частных дисперсий будут:

$$C_A = C'_A \cdot k = 15,36 \cdot 0,906 = 13,916;$$
$$C_B = C'_B \cdot k = 34,60 \cdot 0,906 = 31,347;$$
$$C_C = C'_C \cdot k = 9,4 \cdot 0,906 = 8,516;$$
$$C_{AB} = C'_{AB} \cdot k = 0,28 \cdot 0,906 = 0,253;$$
$$C_{AC} = C'_{AC} \cdot k = 2,52 \cdot 0,906 = 2,283;$$
$$C_{BC} = C'_{BC} \cdot k = 0,24 \cdot 0,906 = 0,217;$$
$$C_{ABC} = C'_{ABC} \cdot k = 4,8 \cdot 0,906 = 4,348,$$

откуда:

$$\Sigma C_{частн} = \Sigma C'_{частн} \cdot k = 60,88 \text{ (или округленно } 60,90).$$

Таблица 43

Сводная таблица дисперсионного анализа (к примеру на стр. 190)

	A	B	C	AB	AC
C	13,916	31,347	8,516	0,253	2,283
$\eta^2 = \frac{C_{фактор}}{C_{общая}}$	0,0946	0,213	0,058	0,002	0,016
ν	1	1	1	1	1
$\sigma^2 = \frac{C}{\nu}$	13,916	31,347	8,516	0,253	2,283
$F = \frac{\sigma^2_{фактор}}{\sigma^2_Z}$	5,33	12,01	3,262	0,098	0,87
F (табличное)	13,2 7,5 4,1	13,2 7,5 4,1	13,2 7,5 4,1	13,2 7,5 4,1	13,2 7,5 4,1

Таким образом, сумма корректированных частных дисперсий оказалась равной той истинной общефакториальной дисперсии, которая была получена по формуле:

$$C_x = \Sigma h_x - H.$$

После получения корректированных частных дисперсий и дисперсий C_x, C_z и C_y остается провести обычные расчёты по завершению дисперсионного анализа в сводной таблице анализа (см. табл. 43).

Проведенный анализ показал, что в общей дисперсии доля влияния организованных и неорганизованных факторов на плодовитость мышей следующая:

1) степень резистентности (A) влияет в объёме 9,46%;
2) величина дозы облучения (B) влияет в объёме 21,3%;
3) возраст мышей при облучении влияет в объёме 5,8%;

Продолжение табл. 43

	BC	ABC	X	Z	y
C	0,217	4,348	60,9	86,2	147,1
$\eta_i^2 = \frac{C_{фактор}}{C_{общая}}$	0,001	0,029	0,414	0,585	1,0
ν	1	1	7	33	39
$\sigma^2 = \frac{C}{\nu}$	0,217	4,348	8,7	2,61	—
$F = \frac{\sigma^2_{фактор}}{\sigma^2_Z}$	0,084	1,67	3,33	—	—
F (табличное)	13,2 7,5 4,1	13,2 7,5 4,1	4,8 3,2 2,3	— — —	— — —

4) влияние сочетания факторов (*AB, AC, BC, ABC*) при отсутствии достоверности было не велико (1—2%);

5) общее влияние всех факторов и их сочетаний η_x^2 составило 41,4% при полной достоверности;

6) влияние неучтенных факторов η_z^2 было велико и составило 58,5%.

Из разобранного примера трехфакторного комплекса видно, что усложнение и объём расчётной работы значительно увеличиваются с увеличением числа факторов, включаемых в комплекс. Поэтому вычислительные работы в усложненных комплексах требуют уже применения счетных машин.

В. Дисперсионный анализ качественных признаков в однофакторных, двухфакторных и трехфакторных комплексах

Дисперсионный анализ качественных признаков не отличается существенно от дисперсионного анализа количественных признаков. Для качественных признаков составляются таблицы статистического комплекса с внесением в них показателей объектов, отобранных из генеральной совокупности по принципу случайной выборки. В таких комплексах распределение частот может быть равномерным, пропорциональным и неравномерным, что и сказывается известным образом на ходе обработки комплекса.

Дисперсия качественных признаков в общей форме выражается следующим образом:

$$C = \Sigma n \cdot (V - M)^2.$$

Так как сумма квадратов отклонения вариантов от средней арифметической представляет собой квадрат среднего квадратического отклонения σ^2, то, следовательно, формула дисперсии может быть дана и в таком виде:

$$C = n\sigma^2.$$

Уже указывалось (стр. 48), что среднее квадратическое отклонение σ для качественных признаков имеет такую формулу:

$$\sigma = \sqrt{p \cdot q},$$

где p и q — доли, указывающие на относительное число особей с тем или другим признаком. Так, если имеется n особей, из которых m особей с одним качественным признаком, то доля их в выборке составляет $p=\frac{m}{n}$, а доля особей с другим качественным признаком равна $q=1-\frac{m}{n}$. Если использовать формулу среднего квадратического отклонения σ для выражения дисперсии качественных признаков, то дисперсия будет иметь такой вид:

$$C = n\cdot\sigma^2 = n\cdot p\cdot q = n\cdot\frac{m}{n}\cdot\left(1-\frac{m}{n}\right),$$

$$\text{или } C = m - \frac{m^2}{n}.$$

что и является основной формулой дисперсии качественных признаков.

В статистическом комплексе качественных признаков построчно записываются значения n_x (число наблюдений по градациям факторов воздействия), затем значения m_x (число наблюдений по каждой градации, указывающее количество особей с наличием изучаемого качественного признака). Следующая строчка дает величины m_x^2 по градациям, далее вычисляются отношения $\frac{m_x^2}{n_x}$, сумма которых по строчке дает значение Σh_x, и последняя строчка комплекса включает данные о величине доли p по каждой градации.

Если комплекс однофакторный, то на указанной обработке и заканчивается подготовительная расчётная работа, в результате которой получаются значения Σn, Σm_x, Σh_x, необходимые для вычисления дисперсий.

Формулы дисперсии однофакторного комплекса следующие:

$$C_X = \Sigma h_X - H;\ C_Z = \Sigma m_X - \Sigma h_X;\ C_y = \Sigma m_X - H.$$

Разберем ход обработки такого комплекса на примере.

Пример. Выяснить влияние сорта на стойкость растений кукурузы к поражению шведской мухой (табл. 44).

По данным табл. вычисляем:

$$H = \frac{(\Sigma m_X)^2}{\Sigma n_X} = \frac{400^2}{550} = \frac{160000}{550} = 290{,}9.$$

Таблица 44

Обработка однофакторного статистического комплекса при качественных признаках

Результативный признак, y	Сорт (А)				Число градаций $l_A=4$
	Кичкасская	Ленинградская	Одесская	Голландская гибридная	Суммарные данные по строчкам
Число растений в группе, n_X	100	150	120	180	$\Sigma n_X = 550$
Число растений, пораженных мухой, m_X	65	105	66	164	$\Sigma m_X = 400$
m_X^2	4225	11025	4356	26806	—
$h_A = \frac{m_X^2}{n_X}$	42,25	73,5	36,3	149,42	$\Sigma h_X = 301,47$
$p = \frac{m_X}{n_X}$	0,65	0,70	0,55	0,91	—

Составляем сводную таблицу дисперсионного анализа (табл. 45).

Таблица 45

Сводная таблица дисперсионного анализа (к примеру на стр. 197)

	x	z	y
C	$\Sigma h_X - H = 301,47 - 290,9 = 10,57$	$\Sigma m_X - \Sigma h_X = 400 - 301,47 = 98,53$	$\Sigma m_X - H = 400 - 290,9 = 109,1$
$\eta_1^2 = \frac{C_{фактор}}{C_y}$	$10,57 : 109,1 = 0,0969$	$98,53 \cdot 109,1 = 0,9031$	1,0
ν	$l_A - 1 = 4 - 1 = 3$	$\Sigma n_X - l_A = 550 - 4 - 546$	$\Sigma n_X - 1 = 550 - 1 = 549$
$\sigma^2 = \frac{C}{\nu}$	$10,57 : 3 = 3,52$	$98,53 : 546 = 0,1806$	—
$F = \sigma^2_{фактор} : \sigma_z^2$	$3.52 : 0,1806 = 19,4$	—	—
F (табличное)	5,7; 3,9; 2,7	—	—

Из сводной таблицы видно, что влияние сортности кукурузы на поражение шведской мухой составляет в общей дисперсии 9,69% при полной достоверности влияния этого фактора.

При двух- и трехфакторных равномерных или пропорциональных комплексах обработка общей таблицы осуществляется построчно, как это делалось в предыдущем примере однофакторного комплекса, т. е. по строкам: n_x, m_x, m_x^2, $h_x = \frac{m_x^2}{n_x}$ и p. Затем в подсобных таблицах обрабатываются данные по каждому из факторов и по каждому из их сочетаний. При этом для каждой градации фактора или сочетаний факторов вычисляются частные значения

$$\Sigma n_A,\ \Sigma m_A,\ \Sigma h_A = \Sigma \frac{m_A^2}{n_A} \text{ и } p_A;\ \ \Sigma n_B,\ \Sigma m_B,\ \Sigma h_B = \Sigma \frac{m_B^2}{n_B} \text{ и } p_B.$$

После такой обработки полученные данные подставляются в формулы. Для двухфакторного комплекса такими формулами будут:

$$C_y = \Sigma m_X - H, \quad C_X = \Sigma h_X - H; \quad C_Z = \Sigma m_X - \Sigma h_X;$$

$$C_A = \Sigma h_A - H, \quad C_B = \Sigma h_B - H; \quad C_{AB} = C_X - C_A - C_B.$$

Последняя формула (C_{AB}) в таком виде может быть использована только в ортогональных, т. е. в равномерных или пропорциональных комплексах.

В табл. 46 приводится расчёт двухфакторного пропорционального комплекса:

$$H = \frac{(\Sigma m_X)^2}{\Sigma n_X} = \frac{150^2}{440} = \frac{22500}{440} = 51,136;$$

$$C_y = \Sigma m_X - H = 150 - 51,136 = 98,864;$$

$$C_X = \Sigma h_X - H = 57,5 - 51,136 = 6,364;$$

$$C_Z = \Sigma m_X - \Sigma h_X = 150 - 57,5 = 92,50.$$

Проведём обработку по каждому фактору и сочетаниям факторов в подсобной табл. 47. Находим частные дисперсии:

$$C_A = \Sigma h_A - H = 56,818 - 51,136 = 5,682;$$

$$C_B = \Sigma h_B - H = 51,75 - 51,136 = 0,614;$$

$$C_{AB} = C_X - C_A - C_B = 6,364 - 5,682 - 0,614 = 0,068.$$

Таблица 46

Обработка двухфакторного пропорционального комплекса при качественных признаках

Результативный признак, x	Фактор A_1		Фактор A_2		Число градаций по $A=2$; по $B=2$
	B_1	B_2	B_1	B_2	Суммарные данные по строчкам
n_x	100	120	100	120	$\Sigma n_x = 440$
m_x	20	30	40	60	$\Sigma m_x = 150$
m_x^2	400	900	1600	3600	—
$h_x = \frac{m_x^2}{n_x}$	$\frac{400}{100} = 4,0$	$\frac{900}{120} = 7,5$	$\frac{1600}{100} = 16,0$	$\frac{3600}{120} = 30$	$\Sigma h_x = 57,5$
$p = \frac{m_x}{n_x}$	0,20	0,25	0,40	0,50	—

Таблица 47

Продолжение обработки статистического комплекса (к примеру из табл. 46)

Факторы и их градации	$\Sigma n_{частн}$	$\Sigma m_{частн}$	$(\Sigma m_{частн})^2$	$h_{частн} = \frac{(\Sigma m_{частн})^2}{\Sigma n_{частн}}$	$p_{частн}$
A_1 A_2	220 220	50 100	2500 10000	2500:220 = 11,363 10000:220 = 45,455	0,22 0,45
	440	—	—	$\Sigma h_A = 56,818$	
B_1 B_2	200 240	60 90	3600 8100	3600:200 = 18,00 8100:240 = 33,75	0,30 0,37
	440			$\Sigma h_B = 51,75$	

Завершаем обработку комплекса составлением сводной табл. 48.

Проведенный анализ по этому примеру показал, что достоверно только суммарное влияние организованных факторов, которое составляет в общей дисперсии 6,4%.

Трехфакторный комплекс для качественных признаков рассчитывается аналогичным образом, формулы лишь дополняются дисперсиями для третьего фактора и для сочетаний, связанных с его наличием.

Перечислим значения строчек при обработке трехфакторного комплекса и его формулы:

а. Значения величин по строчкам основного комплекса n_x, m_x, m_x^2, $h_x = \frac{m_x^2}{n_x}$, p. По этим данным вычисляются Σn_x, Σm_x, $H = \frac{(\Sigma m_x)^2}{\Sigma n_x}$ и Σh_x, а также дисперсии $C_y = \Sigma m_x - H$, $C_x = \Sigma h_x - H$ и $C_z = \Sigma m_x - \Sigma h_x$.

б. Значения величин при обработке подсобных таблиц по факторам и их сочетаниям:

для фактора A:

Σn_A, Σm_A, $(\Sigma m_A)^2$, $\frac{(\Sigma m_A)^2}{\Sigma n_A}$, Σh_A, p_A и дисперсия $C_A = \Sigma h_A - H$;

для фактора B:

Σn_B, Σm_B, $(\Sigma m_B)^2$, $\frac{(\Sigma m_B)^2}{\Sigma n_B}$, Σh_B, p_B и дисперсия $C_B = \Sigma h_B - H$;

для фактора C:

Σn_C, Σm_C, $(\Sigma m_C)^2$, $\frac{(\Sigma m_C)^2}{\Sigma n_C}$, Σh_C, p_C и дисперсия $C_C = \Sigma h_C - H$;

для совместного действия двух факторов AB:

Σn_{AB}, Σm_{AB}, $(\Sigma m_{AB})^2$, $\frac{(\Sigma m_{AB})^2}{\Sigma n_{AB}}$, Σh_{AB}, p_{AB} и дисперсия $C_{AB} = \Sigma h_{AB} - C_A - C_B - H$;

для совместного действия факторов AC:

Σn_{AC}, Σm_{AC}, $(\Sigma m_{AC})^2$, $\frac{(\Sigma m_{AC})^2}{\Sigma n_{AC}}$, Σh_{AC}, p_{AC} и дисперсия $C_{AC} = \Sigma h_{AC} - C_A - C_C - H$;

Таблица 48

Сводная таблица дисперсионного анализа (к примеру из табл. 46)

	A	B	AB	X	Z	y
C	5,682	0,614	0,068	6,364	92,50	98,864
$\eta_i^2 = \frac{C_{фактор}}{C_y}$	0,057	0,0062	0,0006	0,0643	0,935	1,0
ν	2—1=1	2—1=1	1.1=1	4—1=3	440—4=436	440—1=439
$\sigma^2 = \frac{C}{\nu}$	5,682	0,614	0,068	2,121	0,212	—
$F = \frac{\sigma^2_{фактор}}{\sigma^2_Z}$	2,68	2,9	0,3	10,0	—	—
F (табличное)	$\begin{cases} 11,0 \\ 6,7 \\ 3,9 \end{cases}$	$\begin{cases} 11,0 \\ 6,7 \\ 3,9 \end{cases}$	$\begin{cases} 11,0 \\ 6,7 \\ 3,9 \end{cases}$	$\begin{cases} 5,6 \\ 3,8 \\ 2,6 \end{cases}$		

для совместного действия факторов BC:

Σn_{BC}, Σm_{BC}, $(\Sigma m_{BC})^2$, $\frac{(\Sigma m_{BC})^2}{\Sigma n_{BC}}$, Σh_{BC}, p_{BC} и дисперсия $C_{BC} = \Sigma h_{BC} - C_B - C_C - H$;

для совместного действия трех факторов ABC:

$$C_{ABC} = C_x - C_A - C_B - C_C - C_{AB} - C_{AC} - C_{BC}.$$

Остальная обработка осуществляется по форме сводной таблицы дисперсионного комплекса.

Г. Использование дисперсионного анализа для сравнения средних арифметических внутри комплекса и для определения достоверности разности между ними

В ходе дисперсионного анализа можно не только выяснить степень влияния факторов на результативный признак, выраженный относительными величинами η^2, C и σ^2, но и провести *сопоставление абсолютных значений частных средних арифметических*, характеризующих величину признака по каждому фактору, и определить достоверность разности между этими частными средними.

Существенной стороной дисперсионного анализа в такой обработке является то, что он дает более точное представление о достоверности, нежели обычно используемый приём определения достоверности разности между средними по формуле ее ошибки, а именно: $m_D = \sqrt{m_1^2 + m_2^2}$. Это объясняется тем, что ошибки средних, входящие в подкоренную величину, вычисляются по приближенной формуле:

$$m_M = \frac{\sigma}{\sqrt{n}};$$

точная формула ошибки средних следующая:

$$m_M = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}}.$$

При вычислениях достоверности разности между частными средними, полученными при дисперсионном анализе статистического комплекса, пользуются ошибкой, для которой берется среднее квадратическое отклонение σ, получаемое из значения остаточной дисперсии σ_z^2. При этом формула ошибки частной средней выглядит так:

$$m_{M_{частн}} = \frac{\sigma_z}{\sqrt{n_{частн}}}, \quad \text{где} \;\; \sigma_z = \sqrt{\sigma_z^2};$$

$n_{частн}$ — число наблюдений по градациям фактора, для которого вычисляется частная средняя арифметическая.

При вычислении достоверности разности между частными средними $D = M_1 - M_2$ пользуются преобразованной формулой критерия достоверности разности t_D, которая основывается на том, что коэффициенты достоверности F или θ, используемые в дисперсионном анализе (см. стр. 167), связаны с критерием достоверности t следующим равенством:

$$F = \theta = t^2.$$

Поэтому при вычислении критерия достоверности между частными средними дисперсионного комплекса можно перейти от формулы t_D, включающей в себя значение ошибок, к значениям t^2, которое позволит использовать коэффициенты F или θ, применяющиеся в дисперсионном анализе для определения достоверности.

Для этого возведем в квадрат обе части равенства $t_D = \frac{D}{m_D}$, получим $t_D^2 = \frac{D^2}{m_D^2}$. Подставим вместо величины m_D^2 ее значение, выраженное через $m_D = \sqrt{m_1^2 + m_2^2} = \sqrt{\frac{\sigma_z^2}{n_1} + \frac{\sigma_z^2}{n_2}}$; получим;

$$t_D^2 = \frac{D^2}{\left(\frac{\sigma_z^2}{n_1} + \frac{\sigma_z^2}{n_2}\right)},$$

где σ_z^2 — величина остаточной дисперсии данного комплекса;

n_1 и n_2 — число наблюдений по каждому фактору, для которых вычисляется t_D.

Для удобства вычисления можно выразить знаменатель дроби таким образом:

$$\frac{\sigma_z^2}{n_1} + \frac{\sigma_z^2}{n_2} = \sigma_z \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \sigma_z^2 \cdot \frac{n_1 + n_2}{n_1 \cdot n_2},$$

и тогда критерий достоверности разности принимает следующее выражение:

$$F_D = \theta_D = t_D^2 = \frac{D^2}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2}$$

при степенях свободы $\nu_1 = 1$ (по двум средним сравниваемых факторов, т. е. $l - 1 = 1$) и $\nu_2 = \nu_z$.

Далее сравнивается вычисленное значение F с табличным значением F при указанных степенях свободы. Если вычисленное значение F больше табличного значения F, то разница между сравниваемыми средними дисперсионного комплекса достоверна.

Разберем решение такого рода задачи на примере, приведенном на стр. 167 в табл. 33. Здесь рассматривается влияние трех типов рациона (A_1 — рацион с нормальным содержанием белка, A_2 — рацион, обогащенный белком, A_3 — рацион с минимальным содержанием белка) на плодовитость мышей в зависимости от того, получали ли эти рационы самки (B_1) или самцы (B_2), поступающие на размножение.

Частные средние по факторам и градациям из табл. 33 оказались следующими (табл. 49).

Таблица 49

Определение достоверности разности между частными средними арифметическими статистического комплекса

Фактор	Рацион с нормальным содержанием белка, A_1	Рацион, обогащенный белком, A_2	Рацион с минимальным содержанием белка, A_3
		Самки, B_1	
Средняя плодовитость по группам, $M_{частн}$	$\frac{6+10}{2} = 8{,}0$	$\frac{8+11}{2} = 9{,}5$	$\frac{5+3}{2} = 4{,}0$
		Самцы, B_2	
Средняя плодовитость по группам, $M_{частн}$	$\frac{7+8}{2} = 7{,}5$	$\frac{9+10}{2} = 9{,}5$	$\frac{4+5}{2} = 4{,}5$

Определим достоверность разности между некоторыми средними. Так, если требуется сравнить, как влияет пониженный рацион (типа A_3) на плодовитость самок и самцов, то надо сравнивать значения $M_{B_1A_1}$ и $M_{B_2A_1}$. Для этого проводятся расчёты, связанные с получением данных для формулы:

$$F = t_D^2 = \frac{D^2}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2}.$$

Из табл. 49 видно, что частные средние имеют такие значения:

$$M_{B_1A_1} = 4{,}0; \quad M_{B_2A_1} = 4{,}5,$$

отсюда $D = 0{,}5$ голов; $n_{B_1A_1} = 2$ и $n_{B_2A_1} = 2$; из таблицы 35 этого комплекса на стр. 173 известно, что $\sigma_z^2 = 2{,}66$, а $\nu_z = 6$.

Найдем:

$$D^2 = 0{,}5^2 = 0{,}25.$$

Подставим имеющиеся данные в формулу критерия достоверности F:

$$F_D = \frac{D^2}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2} = \frac{0{,}25}{2{,}66} \cdot \frac{2 \cdot 2}{2 + 2} = \frac{0{,}25 \cdot 4}{10{,}64} = 0{,}094.$$

В приложении находим табличные значения F при $\nu_1 = 1$ и $\nu_2 = \nu_z = 6$. Они оказались равны при трех градациях вероятности: 35,5; 13,4; 6,0, т. е. больше вычисленного значения F. Таким образом, разность между этими средними не достоверна, а это означает, что в данном примере рацион с пониженным содержанием белка одинаково влияет на самцов и самок в снижении их плодовитости. Если требуется выяснить, как изменяется плодовитость самок, получавших обогащенный рацион (B_1A_2) по сравнению с плодовитостью самок контрольной группы (A_1B_1), то проводится сравнение величин $M_{B_1A_2}$ и $M_{B_1A_1}$, разность которых равна: 9,5—8,0 = 1,5.

Достоверность этой разности определяется аналогично тому, как это сделано для предыдущего сопоставления. Проведем расчёты для сопоставления M по градации B_1A_2 и B_1A_1:

$$M_{B_1A_2} - M_{B_1A_1} = D = 1{,}5; \quad D^2 = 2{,}25;$$

$$n_1 = 2; \quad n_2 = 2; \quad \sigma_Z^2 = 2{,}66 \text{ и } \nu_Z = 6.$$

Находим:

$$F = \frac{D^2}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2} = \frac{2,25}{2,66} \cdot \frac{2 \cdot 2}{2+2} = \frac{9,0}{10,64} = 0,84.$$

Табличные значения F при $\nu_1 = 1$ и $\nu_2 = \nu_z = 6$ равны соответственно 35,5, 13,4 и 6,0, следовательно, полученная разность в 1,5 мышонка между этими группами оказалась также недостоверной.

Вывод: прибавка белка в рационе сверх нормы не стимулировала плодовитость самок.

В тех случаях, когда имеет место неравномерный комплекс, обработка которого ведется методом Ю. Л. Поморского, рассмотренным на стр. 181 и в табл. 37, когда средние получаются не путем деления ΣV на n, а путем деления суммы средних по градациям на число этих средних i: т. е. $\frac{\Sigma M_x}{i}$, определение достоверности разности осуществляется по несколько иной формуле, а именно:

$$F = \frac{D^2}{\sigma_Z^2} \cdot \frac{i^2}{\Sigma \frac{1}{n_1} + \Sigma \frac{1}{n_2}},$$

где D^2 — квадрат разности между значениями M_1 и M_2;
σ_z^2 — остаточная дисперсия для всего комплекса;
i — число частных средних M_x, из которых получена каждая сравниваемая средняя M_1 и M_2;
n_1 и n_2 — число наблюдений в градациях, по которым получены частные средние M_x.

Примером такой обработки служит двухфакторный комплекс, приведенный на стр. 184 в табл. 38, где частные средние M по градациям факторов, т. е. M_{A_1} и M_{A_2} вычислены обычным путем, т. е. делением ΣV на n. Если потребуется выяснить на матках разных пород, как влияет тип кормления на плодовитость и определить достоверность разности между M_{A_1} и M_{A_2}, то можно пользоваться рассмотренной ранее формулой:

$$F = \frac{D^2}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2}.$$

Но в этом же комплексе имеются и частные средние, вычисленные по градациям каждого фактора отдельно, вне зависимости от градаций другого фактора, по элементарным средним ΣM_x, деленным на число этих элементарных средних в градации i по формуле:

$$M_{частн} = \frac{\Sigma M_x}{i}$$

(см. табл. 39, стр. 186).

Поэтому определение достоверности разности между этими средними (т. е. между значениями M_{A_1}, M_{A_2}, M_{B_1}, M_{B_2}) должно выполняться по другой формуле, упомянутой на стр. 207, а *именно*:

$$F = \frac{D^2}{\sigma_Z^2} \cdot \frac{i^2}{\Sigma \frac{1}{n_1} + \Sigma \frac{1}{n_2}}.$$

Разберем этот случай определения достоверности по данным табл. 38.

В этой таблице выясняется влияние типа кормления (A_1 и A_2) и породы свиноматок (B_1 и B_2) на их потенциальную плодовитость, определяемую числом желтых тел яичника.

Если бы требовалось производить сопоставление средних по признаку x, т. е. сопоставлять значения $M_{A_1B_1}$, $M_{A_1B_2}$ и $M_{A_2B_1}$, $M_{A_2B_2}$, то достоверность определялась бы по разобранной формуле:

$$F = \frac{D^2}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2}.$$

Например, разница в плодовитости, обусловленная породностью свиноматок B_1 и B_2, *получавших один и тот же* рацион (типа A_1), оказалась следующей:

$$M_{A_1B_1} - M_{A_1B_2} = 13,25 - 18,33 = -5,08,$$

т. е. матки породы беркшир имели на 5,08 желтых тел меньше, чем матки породы крупной белой. В этом случае: $n_1 = 4$ и $n_2 = 3$, а $\sigma_z^2 = 5,727$ при $\nu_z = 11$ (*см. сводную табл. 40 анализа на стр. 187*).

Отсюда:

$$F = \frac{(-5,08)^2}{5,727} \cdot \frac{4 \cdot 3}{4 + 3} = \frac{25,8 \cdot 12}{5,727 \cdot 7} = \frac{309,6}{40,09} = 7,72.$$

Табличные значения F при $\nu_1 = 1$ и $\nu_2 = \nu_z = 11$ равны соответственно 19,7; 9,7; 4,8.

Таким образом, породная разница в потенциальной плодовитости маток достоверна (при $P=0{,}95$).

Частные средние показателя плодовитости, полученные от деления ΣM_x на i, были следующими (табл. 39):
$M_{A_1}=15{,}79$ (при концентратном типе кормления);
$M_{A_2}=17{,}56$ (при полуконцентратном типе кормления);
$M_{B_1}=14{,}79$ (у свиноматок беркширской породы);
$M_{B_2}=18{,}56$ (у свиноматок крупной белой породы).

Пусть требуется выяснить, как влияет тип кормления на потенциальную плодовитость без учета породности маток, т. е. сравнить величины M_{A_1} и M_{A_2}.

Находим:

$$D = M_{A_1} - M_{A_2} = 15{,}79 - 17{,}56 = -1{,}77,$$

откуда:

$$D^2 = (-1{,}77)^2 = 3{,}1329; \quad \sigma_Z^2 = 5{,}727; \; \nu_Z = 11; \quad \nu_1 = 1; \quad i = 2.$$

Число частот по градациям каждого фактора, по которым было найдено значение ΣM_x для A_1 и значение ΣM_x для A_2, было следующим: для A_1 число частот n_1 по градациям B было равно 4_{B_1} и 3_{B_2}, а для A_2 число частот n_2 было равно 3_{B_1} и 5_{B_2}. Находим сумму обратных величин $\Sigma \frac{1}{n_1}$ и $\Sigma \frac{1}{n_2}$ для этих частот:

$$\Sigma \frac{1}{n_1} = \frac{1}{4} + \frac{1}{3} = 0{,}25 + 0{,}33 = 0{,}58$$

и

$$\Sigma \frac{1}{n_2} = \frac{1}{3} + \frac{1}{5} = 0{,}33 + 0{,}20 = 0{,}53;$$

откуда:

$$\Sigma \frac{1}{n_1} + \Sigma \frac{1}{n_2} = 0{,}58 + 0{,}53 = 1{,}11.$$

Подставляем необходимые величины в формулу F:

$$F = \frac{D^2}{\sigma_Z^2} \cdot \frac{i^2}{\Sigma \frac{1}{n_1} + \Sigma \frac{1}{n_2}} = \frac{3{,}1329}{5{,}727} \cdot \frac{2^2}{1{,}11} = \frac{12{,}5316}{6{,}3569} = 1{,}97.$$

Табличные значения F при $\nu_1=1$ и $\nu_2=\nu_z=11$ при трех уровнях вероятности равны соответственно 19,7; 9,7; 4,8. Таким образом, разница в плодовитости, обусловленная типами рациона, не достоверна.

Так как недостоверность разности может возникать в результате малого числа наблюдений, включенных в комплекс, возникает необходимость определения численности выборки, которая может дать достоверные результаты при заданном уровне вероятности.

Этот вопрос для статистических рядов рассматривался нами ранее на стр. 115, 116; там указывалось, что необходимая численность выборки для получения достоверных результатов может быть установлена по таблице больших чисел, с помощью номограммы или по формуле.

Для статистических комплексов при дисперсионном анализе численность выборки, необходимая для получения достоверных данных, определяется по следующей формуле:

$$n' = \frac{n + n \cdot (F_1 : F)}{2} = \frac{n}{2} \cdot \left(1 + \frac{F_1}{F} \right),$$

где n' — искомая достаточная численность выборки;

n — численность выборки в комплексе, при обработке которого получены недостоверные результаты;

F — коэффициент достоверности, полученный для данного комплекса;

F_1 — желаемая величина коэффициента достоверности при определенном уровне вероятности P, который даст достоверные результаты при степенях свободы ν_1 и ν_2, имевших место в комплексе с недостоверными результатами.

Пример. В разобранном на стр. 209 примере оказалось, что разница в плодовитости свиноматок, вызываемая разным типом их кормления, оказалась недостоверной. Данный комплекс имел: $n=15$, $F=1{,}97$, $\nu_1=1$ и $\nu_2=\nu_z=11$.

Определить, какое число наблюдений в комплексе могло дать достоверную разницу при $F'=4{,}8$ (по таблице при $P=0{,}95$).

Подставляем в формулу соответствующие значения:

$$n' = \frac{15}{2} \cdot \left(1 + \frac{4{,}8}{1{,}97}\right) = 7{,}5 \cdot \left(1 + \frac{4{,}8}{1{,}97}\right) = 7{,}5 \cdot 3{,}43 = 25{,}72.$$

Следовательно, округляя n', при 25—26 наблюдениях в комплексе можно получить достоверные данные о различиях в сравниваемых частных средних данного примера при уровне вероятности $P=0{,}95$.

Важным вопросом при анализе вариантов дисперсионного комплекса является вопрос о том, случайно или не случайно отдельные варианты по своей абсолютной величине значительно отклоняются в сторону увеличения или уменьшения от других вариантов данного комплекса. Наличие таких единичных, с сильным отклонением, вариантов может быть или в результате методических ошибок или обусловлено какими-то неизвестными факторами, закономерно вызвавшими данное отклонение на фоне обычной вариабильности данного признака.

Такое явление называется «выпадом».

Для выяснения случайности или неслучайности такого отклонения пользуются величиной F, вычисленной на основании следующей формулы:

$$F = \frac{(V_i - M_i)^2}{(\sigma_Z^2)'} \cdot \frac{n_i - 1}{n_i} \text{ при } \nu_1 = 1 \text{ и } \nu_2 = \nu_Z = 1,$$

где V_i — величина отклоняющегося (проверяемого) варианта;

M_i — средняя арифметическая, получаемая из вариантов той же градации, где находился отклоняющийся вариант V_i и после его исключения;

n_i — число вариантов в той градации, где находится проверяемый вариант V_i;

$(\sigma_z^2)'$ — остаточная дисперсия, получаемая по оставшейся части вариантов комплекса после удаления отклоняющегося варианта, которая может быть получена расчётным путем следующим образом:

$$(\sigma_Z^2)' = \frac{C_Z'}{\nu_Z - 1}.$$

Значение новой остаточной дисперсии C_z' можно также вычислить по формуле:

$$C_Z' = C_Z - \frac{n_i}{n_i - 1} \cdot (V_i - M_i)^2,$$

где C_z — остаточная дисперсия по всему комплексу с включением и отклоняющегося варианта V_i;

M_i — средняя по той градации, где находится отклоняющийся вариант; она вычисляется так: $M_i = \frac{\Sigma V_i}{n_i}$.

Если вычисленное по приведенной формуле значение F больше табличного значения F, то разность между величиной проверяемого варианта V_i и остальными вариантами комплекса достоверна. В этом случае такой вариант, как не случайно попавший в комплекс, должен быть устранен из комплекса при обработке последнего.

После исключения варианта, являющегося «выпадом», обработку уточненного по вариантам комплекса делают заново и получают новые значения C, η, σ и F, отражающие закономерности варьирования в корректированном комплексе.

Определение достоверности разности между двумя средними можно делать также для двух различных вариационных рядов, пользуясь принципами достоверности F дисперсионного анализа по формуле:

$$F = \frac{D^2}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2} \text{ при } \nu_1 = 1 \text{ и } \nu_2 = n_1 + n_2 - 2,$$

где

$$\sigma_Z^2 = \frac{C_Z}{n_1 + n_2 - 2}; \; C_Z = \Sigma V^2 - \Sigma h_X;$$

$$\Sigma V^2 = \Sigma V_1^2 + \Sigma V_2^2; \; \Sigma h_X = \frac{(\Sigma V_1)^2}{n_1} + \frac{(\Sigma V_2)^2}{n_2}.$$

Чтобы воспользоваться таким методом, необходимо объединить оба вариационных ряда в своего рода статистический комплекс. При этом могут быть разные формы вариационных рядов, а именно: с одинаковой и различной величиной класса. В зависимости от этого техника обработки несколько меняется.

Разберем пример для двух рядов с одинаковым значением величины класса (κ).

Пример. Определить достоверность разности между средним числом завязей на облученных и необлученных растениях томатов.

Вариационный ряд для необлученных растений:

Классы по числу завязей	20—29	30—39	40—49	50—59	60—69
Частоты, p	2	8	50	30	10

Вариационный ряд для облученных растений:

Классы по числу завязей	0—9	10—19	20—29	30—39	40—49
Частоты, p	10	30	40	10	10

Для обоих рядов величина класса $k=10$.

Объединим оба вариационных ряда в виде однофакторного комплекса (см. табл. 50).

Находим для каждого ряда Σxp, $(\Sigma xp)^2$, $h=\frac{(\Sigma xp)^2}{n}$, Σx^2p и M, значения которых необходимы для вычисления дисперсий и показателя достоверности F (табл. 51).

Из данных таблицы вычисляем ΣV^2, Σh_x и C_z:

$$\Sigma V^2 = \Sigma x^2p_1 + \Sigma x^2p_2 = 245300 + 64500 = 309800;$$

$$\Sigma h_X = 238144 + 52900 = 291044;$$

$$C_Z = \Sigma V^2 - \Sigma h_X = 309800 - 291044 = 18756;$$

$$\nu_Z = n_1 + n_2 - 2 = 100 + 100 - 2 = 198;$$

$$\sigma_Z^2 = \frac{C_Z}{\nu_Z} = \frac{18756}{198} = 94{,}72.$$

Подставляем необходимые данные в формулу F и находим достоверность разности между средним числом завязей у растений обоих рядов:

$$F = \frac{(M_1 - M_2)^2}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2} = \frac{(48{,}8 - 23)^2}{94{,}72} \cdot \frac{100 \cdot 100}{100 + 100} =$$

$$= \frac{(25{,}8)^2 \cdot 10000}{18944} = \frac{665{,}64 \cdot 10000}{18944} = 351{,}3.$$

Табличные значения F при $\nu_1 = 1$ и $\nu_2 = \nu_z = 198$ равны соответственно 11,2; 6,8; 3,9. Следовательно, разность между средними обоих рядов достоверна при всех уровнях вероятности.

Так как разобранный пример характерен громоздкими величинами x и p, то его данные могли бы быть обработаны с использованием условного отклонения a. Эта обработка дана в табл. 52

Проведем дальнейшие вычисления величин, необходимых при отыскании σ_z^2 и F (табл. 53).

$$\Sigma h = 190{,}44 + 144{,}0 = 334{,}44; \quad \Sigma\Sigma pa^2 = 262 + 260 = 522;$$

$$C_Z = \Sigma\Sigma pa^2 - \Sigma h = 522 - 334{,}44 = 187{,}56;$$

$$\nu_Z = n_1 + n_2 - 2 = 100 + 100 - 2 = 198;$$

$$\sigma_Z^2 = \frac{C_Z}{\nu_Z} = \frac{187{,}56}{198} = 0{,}9472.$$

Таблица 50

Определение достоверности разности между средними арифметическими двух вариационных рядов при одинаковых размерах класса

x	Середина класса x	Частоты по рядам		xp		x^2p	
		p_1	p_2	1	2	1	2
0—9	5	—	10	—	50	—	250
10—19	15	—	30	—	450	—	6750
20—29	25	2	40	50	1000	1250	25000
30—39	35	8	10	280	350	9800	12250
40—49	45	50	10	2250	450	101250	20250
50—59	55	30	—	1650	—	90750	—
60—69	65	10	—	650	—	42250	—
		$\Sigma p_1=100$	$\Sigma p_2=100$	$\Sigma xp_1=4880$	$\Sigma xp_2=2300$	$\Sigma x^2p_1=245300$	$\Sigma x^2p_2=64500$

Таблица 51

Продолжение обработки данных (к примеру из табл. 50)

	n	Σxp	$(\Sigma xp)^2$	$h=(\Sigma xp)^2:n$	Σx^2p	$M=\Sigma xp:n$
1-й ряд	100	4880	$4880^2=23814400$	$\frac{23814400}{100}=238144$	245300	4880:100=48,80
2-й ряд	100	2300	$2300^2=5290000$	$\frac{5290000}{100}=52900$	64500	2300:100=23,0

Таблица 52

Вычисление достоверности разности между средними арифметическими двух вариационных рядов при большом числе наблюдений

Середина класса x	Ряды		a	$p_1 a$	$p_2 a$	$p_1 \cdot a^2$	$p_2 a^2$
	p_1	p_2					
5	—	10	—3	—	—30	—	90
15	—	30	—2	—	—60	—	120
25	2	40	—1	—2	—40	2	40
35	8	10	0	0	0	0	0
45	50	10	1	50	10	50	10
55	30	—	2	60	—	120	—
65	10	—	3	30	—	90	—
	$n_1=\Sigma p_1=100$	$n_2=\Sigma p_2=100$	—	$\Sigma p_1 \cdot a=138$	$\Sigma p_2 \cdot a=-120$	$\Sigma p_1 \cdot a^2=262$	$\Sigma p_2 \cdot a^2=260$

Таблица 53

Продолжение обработки данных (к примеру из табл. 52)

Ряд	$n=\Sigma p$	Σpa	Σpa^2	$(\Sigma pa)^2$	$h=\frac{(\Sigma pa)^2}{n}$	$b=\frac{\Sigma pa}{n}$
1-й	100	138	262	$138^2=19044$	$\frac{19044}{100}=190,44$	1,38
2-й	100	—120	260	$(-120)^2=14400$	$\frac{14400}{100}=144,0$	—1,2

Находим значения средних:

$$M_1 = A + k \cdot b_1 = 35 + 10 \cdot 1{,}38 = 35 + 1{,}38 = 48{,}8;$$
$$M_2 = A + k \cdot b_2 = 35 + 10 \cdot (-1{,}20) = 35 - 12 = 23{,}0.$$

Вычисляем достоверность разности между M_1 и M_2, выражая ее разностью в показателях $(b_1 - b_2)$.

$$F = \frac{(b_1 - b_2)^2}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2} = \frac{[1{,}38 - (-1{,}2)]^2}{0{,}9472} \cdot \frac{100 \cdot 100}{100 + 100} =$$
$$= \frac{(1{,}38 + 1{,}2)^2}{0{,}9472} \cdot \frac{10000}{200} = \frac{2{,}58^2 \cdot 10000}{189{,}44} =$$
$$= \frac{6{,}6564 \cdot 10000}{189{,}44} = \frac{66564}{189{,}44} = 351{,}3.$$

Таким образом, достоверность разности между средними, вычисленная обоими способами, оказалась одинаковой.

Если вариационные ряды имеют разное значение величины классов (κ), то обработка их осуществляется раздельно для получения необходимых промежуточных данных, что показано в табл. 54.

Если обработка рядов проведена с использованием условного отклонения a, то остаточная дисперсия выражается следующим образом: $C_z = k_1^2 \cdot C_1 + k_2^2 \cdot C_2$, т. е. она равна сумме произведений квадратов классовых промежутков на свою дисперсию, вычисленных для каждого ряда.

Число степеней свободы для C_z равно: $\nu_z = n_1 + n_2 - 2$; в приведенном примере: $\nu_z = 100 + 100 - 2 = 198$.

Путем вычисления находим:

$$C_Z = k_1^2 \cdot C_1 + k_2^2 \cdot C_2 = 5^2 \cdot 129 + 3^2 \cdot 92{,}75 =$$
$$= 3225 + 834{,}75 = 4059{,}75$$

и

$$\sigma_Z^2 = \frac{4059{,}75}{198} = 20{,}503.$$

Определяем теперь достоверность разности между средними M_1 и M_2:

$$F = \frac{(15{,}5 - 12{,}45)^2}{20{,}503} \cdot \frac{100 \cdot 100}{100 + 100} = \frac{3{,}05^2}{20{,}503} \cdot \frac{10000}{200} =$$
$$= \frac{93025}{4100{,}6} = 22{,}6.$$

Таблица 54

Вычисление достоверности разности между средними арифметическими двух вариационных рядов при разных значениях величины классов

x_1	p_1	a_1	$p_1 a_1$	$p_1 a_1^2$	x_2	p_2	a_2	$p_2 a_2$	$p_2 a_2^2$
5 10	10 20	—2 —1	—20 —20	40 20	6 9	5 15	—2 —1	—10 —15	20 15
15	30	0	0		12	40	0	0	0
20 25	30 10	1 2	30 20	30 40	15 18	20 20	1 2	20 20	20 40
	$\Sigma p_1=100$	—	$\Sigma p_1 a_1=10$	$\Sigma p_1 a_1^2=130$	—	$\Sigma p_2=100$	—	$\Sigma p_2 a=15$	$\Sigma p_2 a_2^2=95$

Таблица 55

Продолжение обработки данных (к примеру из табл. 54)

Ряд	n	Σpa	Σpa^2	$(\Sigma pa)^2$	$h=\frac{(\Sigma pa)^2}{n}$	$b=\frac{\Sigma pa}{n}$	$M=A+\kappa b$	$C=\Sigma pa^2-\Sigma h$
1-й	100	10	130	$10^2=100$	$\frac{100}{100}=1$	$\frac{10}{100}=0,1$	$M_1=15+5\cdot 0,1=$ $=15,5$	$130-1=129$
2-й	100	15	95	$15^2=225$	$\frac{225}{100}=2,25$	$\frac{15}{100}=0,15$	$M_2=12+3\cdot 0,15=$ $=12,45$	$95-2,25=92,75$

Табличные значения F при $\nu_1=1$ и $\nu_2=\nu_z=198$ равны соответственно 11,2; 6,8; 3,9, следовательно, разность между средними двух рядов достоверна при всех уровнях вероятности.

Дисперсионный анализ удобно применять при определении достоверности разности между средними арифметическими двух рядов, если сведения о самих рядах отсутствуют, а имеются только данные о величине значений n, M и σ.

Формула достоверности разности между средними в этом случае используется та же самая, что и в последнем рассмотренном случае, т. е.:

$$F_D = \frac{(M_1 - M_2)^2}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2}.$$

Остаточная дисперсия определяется из выражения:

$$\sigma_Z^2 = \frac{C_1 + C_2}{n_1 + n_2 - 2};$$

где C_1 и C_2 вычисляются из формулы: $\sigma^2 = \frac{C}{\nu}$, откуда $C = \sigma^2 \nu$; $C_z = C_1 + C_2$.

В этом случае ход обработки данных для вычисления достоверности разности между средними при отсутствии рядов выглядит так (табл. 56).

Отсюда:

$$F_D = \frac{(M_1 - M_2)^2}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2} = \frac{(30-20)^2}{3,5} \cdot \frac{100 \cdot 100}{100+100} =$$

$$= \frac{10^2 \cdot 100 \cdot 100}{3,5 \cdot 200} = \frac{1.000.000}{700} = 1428,57.$$

Табличные значения F при $\nu_1=1$ и $\nu_2=\nu_z=198$ равны соответственно 11,2; 6,8; 3,9, что указывает на полную достоверность разности.

Дисперсионный критерий достоверности F используется для определения достоверности разности между средними не только в тех случаях, когда признак имеет количественное выражение, а и для тех данных, когда признак имеет качественное выражение и вместо сравнения средних арифметических производится сравнение долей объектов, имеющих данный признак.

Формула при этом остается такой же, как и для количественных признаков, а именно:

$$F_D = \frac{D^2}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2} \text{ при } \nu_1 = 1 \text{ и } \nu_2 = \nu_Z.$$

Таблица 56

Вычисление достоверности разности между средними арифметическими при отсутствии данных о самих вариационных рядах

Вариационный ряд	Известно				
	n	M	σ	σ^2	$C=\nu\cdot\sigma^2$
1-й	100	30	2,0	4,0	$(100-1)\cdot 4=396$
2-й	100	20	3,0	9,0	$(100-1)\cdot 3=297$

Продолжение табл. 56

Вариационный ряд	Вычисляется		
	$C_Z=C_1+C_2$	$\nu_Z=n_1+n_2-2$	$\sigma_Z^2=\frac{C_Z}{\nu_Z}$
1-й 2-й	$C=396+297=693$	$100+100-2=198$	$\frac{693}{198}=3,5$

Разность D при этом определяется, как разность между сравниваемыми долями.

Разберем ход вычисления достоверности разности между долями.

Пример. Облучение мышей лучами рентгена вызвало появление уродств у потомства, доля которых в группе мышей, принадлежащих резистентной линии, составляла $^1/_{10}$, а нерезистентной линии $^3/_5$.

Определить, достоверна ли разность в появлении уродливого потомства резистентной линии по сравнению с нерезистентной. В группе резистентных мышей было получено 90 нормальных и 10 уродливых мышат, а в нерезистентной группе получено 60 нормальных и 90 уродливых мышат.

Сравниваемые доли признака были следующими: резистентная группа имела долю уродливых мышат $p_1 = {}^{10}/_{100} = {}^1/_{10}$, в нерезистентной группе доля уродств у мышат была $p_2 = {}^{90}/_{150} = {}^{3}/_{5}$.

Число наблюдений по группам соответственно равно $n_1 = 100$ и $n_2 = 150$. Осуществляем по этим данным вычисления дисперсионного комплекса (табл. 57).

Таблица 57

Вычисление достоверности разности между долями при качественных признаках

Группа животных	n	m	m^2	$h_X = \frac{m^2}{n}$	$p = \frac{m}{n}$
резистентная . . .	100	10	100	$\frac{100}{100} = 1{,}0$	$p_1 = \frac{1}{10}$
нерезистентная . .	150	90	8100	$\frac{8100}{150} = 54{,}0$	$p_2 = \frac{3}{5}$
	$\Sigma n = 250$	$\Sigma m = 100$	—	$\Sigma h_x = 55{,}0$	—

$$D = p_1 - p_2 = \frac{1}{10} - \frac{3}{5} = -0{,}5;\ D^2 = (-0{,}5)^2 = 0{,}25.$$

Находим остаточную дисперсию C_z:

$$C_Z = \Sigma m - \Sigma h_X = 100 - 55 = 45; \quad \nu_Z = n_1 + n_2 - 2 =$$
$$= 250 - 2 = 248.$$

Отсюда:

$$\sigma_Z^2 = \frac{C_Z}{\nu_Z} = \frac{45}{248} = 0{,}181.$$

После отыскания σ_z^2 вычисляем достоверность разности между долями p_1 и p_2:

$$F_D = \frac{D^2}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2} = \frac{0{,}25}{0{,}181} \cdot \frac{100 \cdot 150}{100 + 150} = \frac{3750}{45{,}25} = 82{,}8.$$

Табличные значения F при $\nu_1 = 1$ и $\nu_2 = \nu_z = 248$ равны соответственно 11,2; 6,8; 3,9. Так как вычисленное значение превышает табличное значение F, то разность в долях рождения уродливого потомства у двух генетических линий мышей оказалась достоверной.

Заключение

В изложенном практическом курсе по биометрии приводится основная система статистической обработки материала. Рассмотрены методы вычисления средних статистических величин, даны приёмы определения изменчивости количественных и качественных признаков и графическое изображение нормального, асимметрического, эксцессивного и трансгрессивного варьирований. Разобраны различные случаи и способы измерения величины связей между двумя признаками и факторами как для количественных, так и для качественных показателей.

Подробно изложен вопрос о статистических ошибках, критерии достоверности статистических величин для большой и малой выборок. Приведен метод хи-квадрат как приём сопоставления теоретических и эмпирических частот. Разбирается метод дисперсионного анализа для различных типов статистических комплексов и использование критерия достоверности для различных случаев определения достоверности разности между средними арифметическими.

Вместе с тем в книге не изложены некоторые новые методы статистического анализа, которые за последние годы все больше привлекают к себе биологов и которые можно почерпнуть в специальной литературе после освоения основного курса.

Для самостоятельных упражнений по изложенным вопросам рекомендуется использовать пособие П. Ф. Рокицкого «Основы вариационной статистики для биологов», Минск, 1961, в котором приводится значительное количество задач по различным разделам курса. По дисперсионному методу рекомендуется книга Н. А. Плохинского «Дисперсионный анализ», Новосибирск, 1960, где достаточно ясно на примерах разбираются методы дисперсионного анализа.

Для более углубленного использования статистической обработки в научных работах целесообразно пользоваться книгой Дж. У. Снедекора «Статистические методы в применении к исследованиям в сельском хозяйстве и биологии», Сельхозгиз, М., 1961.

Таблица значений коэффициента F

(стандартные отношения девиат $\sigma_x^2 : \sigma_z^2$, соответствующие трем степеням вероятности достоверного разнообразия:
$P_1 = 0{,}95$, $P_2 = 0{,}99$, $P_3 = 0{,}999$ при степенях свободы $\nu_1 = \nu_x$ для σ_x^2 и $\nu_2 = \nu_z$ для σ_z^2)

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	11	12	ν_1 / ν_2
3	167,5 34,1 10,1	148,5 30,8 9,6	141,1 29,5 9,3	137,1 28,7 9,1	134,6 28,2 9,0	132,9 27,9 8,9	131,8 27,7 8,9	130,6 27,5 8,8	131,0 27,4 8,9	129,5 27,2 8,8	128,9 27,1 8,8	128,3 27,1 8,7	3
4	74,1 21,2 7,7	61,2 18,8 6,9	56,1 16,7 6,6	53,4 16,0 6,4	51,7 15,5 6,3	50,5 15,2 6,2	49,8 15,0 6,1	49,0 14,8 6,0	48,6 14,7 6,0	48,2 14,5 6,0	47,8 14,5 5,9	47,4 14,4 5,9	4
5	47,0 16,3 6,6	36,6 13,3 5,8	33,2 12,1 5,4	31,1 11,4 5,2	29,8 11,0 5,1	28,8 10,7 5,0	28,2 10,5 4,9	27,6 10,3 4,8	27,3 10,2 4,8	27,0 10,1 4,7	26,7 10,0 4,7	26,4 9,9 4,7	5
6	35,5 13,4 6,0	27,0 10,9 5,1	23,7 9,8 4,8	21,9 9,2 4,5	20,8 8,8 4,4	20,0 8,5 4,3	19,5 8,3 4,2	19,0 8,1 4,1	18,8 8,0 4,1	18,5 7,9 4,1	18,3 7,8 4,0	18,0 7,7 4,0	6
7	29,2 12,3 5,6	21,7 9,6 4,7	18,8 8,5 4,4	17,2 7,9 4,1	16,2 7,5 4,0	15,5 7,2 3,9	15,1 7,0 3,8	14,6 6,8 3,7	14,4 6,7 3,7	14,2 6,6 3,6	13,9 6,5 3,6	13,7 6,5 3,6	7
8	25,4 11,3 5,3	18,5 8,7 4,5	15,8 7,6 4,1	14,4 7,0 3,8	13,5 6 6 3,7	12,9 6,4 3,6	12,5 6,2 3,5	12,0 6,0 3,4	11,8 5,9 3,4	11,6 5,8 3,3	11,4 5,7 3,1	11,2 5,7 3,3	8
9	22,9 10,6 5,1	16,4 8,0 4,3	13,9 7,0 3,9	12,6 6,4 3,6	11,7 6,1 3,5	11,1 5,8 3,4	10,8 5,6 3,3	10,4 5,5 3,2	10,2 5,4 3,2	10,0 5,3 3,1	9,8 5,2 3,1	9,6 5,1 3,1	9
ν_2 / ν_1	1	2	3	4	5	6	7	8	9	10	11	12	ν_2 \ ν_1

ν_2 \ ν_1	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_1 / ν_2
3	127,7 26,9 8,7	127,1 26,8 8,7	125,5 26,7 8,7	125,9 26,6 8,6	125,6 26,5 8,6	125,3 26,4 8,6	125,0 26,4 8,6	124,7 26,3 8,6	124,4 26,2 8,6	124,1 26,2 8,5	123,8 26,1 8,5	123,5 26,1 8,5	12,9 5,8 3,2	3
4	47,0 14,2 5,9	46,6 14,1 5,8	46,2 14,0 5,8	45,8 13,9 5,8	45,6 13,8 5,7	45,4 13,7 5,7	45,2 13,7 5,7	45,0 13,6 5,7	44,7 13,6 5,7	44,5 13,6 5,7	44,3 13,5 5,6	44,1 13,5 5,6	8,6 4,6 2,8	4
5	26,1 9,8 4,6	25,8 9,7 4,6	25,4 9,6 4,6	25,1 9,5 4,5	24,9 9,4 4,5	24,8 9,3 4,5	24,6 9,2 4,4	24,5 9,1 4,4	24,3 9,1 4,4	24,1 9,1 4,4	24,0 9,0 4,4	23,8 9,0 4,4	6,9 4,0 2,6	5
6	17,7 7,6 4,0	17,5 7,5 3,9	17,2 7,4 3,9	16,9 7,3 3,8	16,8 7,2 3,8	16,6 7,1 3,8	16,5 7,1 3,8	16,4 7,0 3,7	16,2 7,0 3,7	16,1 6,9 3,7	15,9 6,9 3,7	15,8 6,9 3,7	6,0 3,7 2,4	6
7	13,5 6,4 3,5	13,2 6,3 3,5	13,0 6,2 3,4	12,7 6,1 3,4	12,6 6,0 3,4	12,5 5,9 3,3	12,3 5,9 3,3	12,2 5,8 3,3	12,1 5,8 3,3	12,0 5,7 3,3	11,8 5,7 3,2	11,7 5,7 3,2	5,3 3,5 2,4	7
8	11,0 5,6 3,2	10,8 5,5 3,2	10,5 5,4 3,2	10,3 5,3 3,1	10,2 5,2 3,1	10,1 5,1 3,1	10,0 5,1 3,0	9,9 5,0 3,0	9,7 5,0 3,0	9,6 4,9 3,0	9,5 4,9 2,9	9,4 4,9 2,9	5,0 3,4 2,3	8
9	9,4 5,0 3,0	9,2 4,9 3,0	8,9 4,8 2,9	8,7 4,7 2,9	8,6 4,6 2,8	8,5 4,6 2,8	8,4 4,5 2,8	8,3 4,5 2,8	8,1 4,4 2,8	8,0 4,4 2,7	7,9 4,3 2,7	7,8 4,3 2,7	4,8 3,3 2,3	9
ν_2 / ν_1	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_2 \ ν_1

Продолжение

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	11	12	ν_1 / ν_2
10	21,0 10,0 5,0	14,9 7,9 4,1	12,3 6,6 3,7	11,3 6,0 3,5	10,5 5,6 3,3	9,9 5,4 3,2	9,6 5,2 3,1	9,2 5,1 3,1	9,0 5,0 3,0	8,9 4,9 3,0	8,7 4,8 2,9	8,5 4,7 2,9	10
11	19,7 9,7 4,8	13,8 7,2 4,0	11,6 6,2 3,6	10,4 5,7 3,4	9,6 5,3 3,2	9,1 5,1 3,1	8,8 4,9 3,0	8,4 4,7 3,0	8,2 4,6 2,9	8,0 4,5 2,9	7,8 4,5 2,8	7,6 4,4 2,8	11
12	18,6 9,3 4,8	12,3 6,9 3,9	10,8 6,0 3,5	9,6 5,4 3,3	8,9 5,1 3,1	8,4 4,8 3,0	8,1 4,7 2,9	7,7 4,5 2,9	7,5 4,4 2,8	7,4 4,3 2,8	7,2 4,2 2,7	7,0 4,2 2,7	12
13	17,8 9,1 4,7	12,3 6,7 3,8	10,2 5,7 3,4	9,1 5,2 3,2	8,4 4,9 3,0	7,9 4,6 2,9	7,6 4,4 2,8	7,2 4,3 2,8	7,0 4,2 2,7	6,9 4,1 2,7	6,7 4,0 2,6	6,5 4,0 2,6	13
14	17,1 8,9 4,6	11,8 6,5 3,7	9,7 5,6 3,3	8,6 5,0 3,1	7,9 4,7 3,0	7,4 4,5 2,9	7,1 4,3 2,8	6,8 4,1 2,7	6,6 4,0 2,7	6,5 3,9 2,6	6,3 3,9 2,6	6,1 3,8 2,5	14
15	16,6 8,7 4,5	11,3 6,4 3,7	9,3 5,4 3,3	8,3 4,9 3,1	7,6 4,6 2,9	7,1 4,3 2,8	6,8 4,1 2,7	6,5 4,0 2,6	6,3 3,9 2,6	6,2 3,8 2,6	6,0 3,7 2,5	5,8 3,7 2,5	15
16	16,1 8,5 4,5	11,0 6,2 3,6	9,0 5,3 3,2	7,9 4,8 3,0	7,3 4,4 2,9	6,8 4,2 2,7	6,5 4,0 2,7	6,2 3,9 2,6	6,1 3,8 2,5	5,9 3,7 2,5	5,8 3,6 2,5	5,6 3,5 2,4	16
ν_2 / ν_1	1	2	3	4	5	6	7	8	9	10	11	12	ν_2 \ ν_1

Продолжение

ν_2 \ ν_1	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_1 / ν_2
10	8,3 4,6 2,9	8,1 4,5 2,8	7,8 4,4 2,8	7,6 4,3 2,7	7,5 4,3 2,7	7,4 4,2 2,7	7,3 4,1 2,6	7,2 4,1 2,6	7,1 4,0 2,6	7,0 4,0 2,6	6,9 3,9 2,6	6,8 3,9 2,5	4,6 3,2 2,2	10
11	7,4 4,3 2,7	7,3 4,2 2,7	7,1 4,1 2,7	6,9 4,0 2,6	6,8 3,9 2,6	6,7 3,9 2,5	6,6 3,8 2,5	6,5 3,7 2,5	6,3 3,7 2,5	6,2 3,7 2,4	6,1 3,6 2,4	6,0 3,6 2,4	4,4 3,1 2,2	11
12	6,8 4,1 2,6	6,7 4,0 2,6	6,5 3,9 2,5	6,3 3,8 2,5	6,2 3,7 2,5	6,1 3,6 2,4	6,0 3,6 2,4	5,9 3,5 2,4	5,7 3,5 2,4	5,6 2,4 2,3	5,5 3,4 2,3	5,4 3,4 2,3	4,3 3,1 2,2	12
13	6,3 3,9 2,6	6,2 3,8 2,5	6,0 3,7 2,5	5,8 3,6 2,4	5,7 3,5 2,4	5,6 3,4 2,3	5,5 3,4 2,3	5,4 3,3 2,3	5,3 3,3 2,3	5,2 3,2 2,2	5,1 3,2 2,2	5,0 3,2 2,2	4,1 3,0 2,2	13
14	5,9 3,7 2,5	5,8 3,6 2,4	5,6 3,5 2,4	5,4 3,4 2,4	5,3 3,3 2,3	5,2 3,3 2,3	5,1 3,2 2,2	5,0 3,1 2,2	4,9 3,1 2,2	4,8 3,1 2,2	4,7 3,0 2,1	4,6 3,0 2,1	4,1 3,0 2,1	14
15	5,6 3,6 2,4	5,5 3,5 2,4	5,3 3,4 2,3	5,1 3,3 2,3	5,0 3,2 2,3	4,9 3,1 2,2	4,8 3,1 2,2	4,7 3,0 2,2	4,6 3,0 2,1	4,5 2,9 2,1	4,4 2,9 2,1	4,3 2,9 2,1	4,1 3,0 2,1	15
16	5,4 3,5 2,4	5,3 3,4 2,3	5,1 3,3 2,3	4,9 3,2 2,2	4,8 3,1 2,2	4,7 3,0 2,2	4,6 3,0 2,1	4,5 2,9 2,1	4,4 2,9 2,1	4,3 2,8 2,0	4,2 2,8 2,0	4,1 2,8 2,0	4,0 2,9 2,1	16
ν_2 / ν_1	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_2 \ ν_1

Продолжение

ν_1 / ν_2	1	2	3	4	5	6	7	8	9	10	11	12	ν_1 / ν_2
17	15,7 8,4 4,5	10,7 6,1 3,6	8,7 5,2 3,2	7,7 4,7 3,0	7,0 4,3 2,8	6,6 4,1 2,7	6,3 3,9 2,6	6,0 3,8 2,6	5,8 3,7 2,5	5,7 3,6 2,5	5,5 3,5 2,4	5,3 3,5 2,4	17
18	15,4 8,3 4,4	10,4 6,0 3,5	8,5 5,1 3,2	7,5 4,6 2,8	6,8 4,2 2,8	6,4 4,0 2,7	6,1 3,8 2,6	5,8 3,7 2,5	5,6 3,6 2,5	5,5 3,5 2,4	5,3 3,4 2,4	5,1 3,4 2,3	18
19	15,1 8,2 4,4	10,2 5,9 3,5	8,3 5,0 3,1	7,3 4,5 2,9	6,6 4,2 2,7	6,2 3,9 2,6	5,9 3,8 2,5	5,6 3,6 2,5	5,5 3,5 2,4	5,3 3,4 2,4	5,2 3,4 2,3	5,0 3,3 2,3	19
20	14,8 8,1 4,3	10,0 5,8 3,5	8,1 4,9 3,1	7,1 4,4 2,9	6,5 4,1 2,7	6,0 3,9 2,6	5,7 3,7 2,5	5,4 3,6 2,4	5,3 3,4 2,4	5,1 3,4 2,3	5,0 3,3 2,3	4,8 3,2 2,3	20
21	14,6 8,0 4,3	9,8 5,8 3,5	7,9 4,9 3,1	7,0 4,4 2,8	6,3 4,0 2,7	5,9 3,8 2,6	5,6 3,6 2,5	5,3 3,5 2,4	5,2 3,4 2,4	5,0 3,3 2,3	4,9 3,2 2,3	4,7 3,2 2,2	21
22	14,4 7,9 4,3	9,6 5,7 3,4	7,7 4,8 3,0	6,8 4,3 2,8	6,2 4,0 2,7	5,8 3,8 2,6	5,5 3,6 2,5	5,2 3,4 2,4	5,1 3,3 2,3	4,9 3,3 2,3	4,8 3,2 2,3	4,6 3,1 2,2	22
23	14,2 7,9 4,3	9,5 5,7 3,4	7,8 4,8 3,0	6,7 4,3 2,8	6,1 4,0 2,6	5,6 3,7 2,5	5,4 3,5 2,4	5,1 3,4 2,4	5,0 3,3 2,3	4,8 3,2 2,3	4,7 3,1 2,2	4,5 3,1 2,2	23
ν_2 / ν_1	1	2	3	4	5	6	7	8	9	10	11	12	ν_2 / ν_1

Продолжение

ν_2 \ ν_1	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_1 / ν_2
17	5,1 3,4 2,3	5,0 3,3 2,3	4,8 3,2 2,2	4,6 3,1 2,2	4,5 3,0 2,2	4,4 2,9 2,1	4,3 2,9 2,1	4,3 2,8 2,0	4,2 2,8 2,0	4,1 2,7 2,0	4,0 2,7 2,0	3,9 2,7 2,0	4,0 2,9 2,1	17
18	5,0 3,3 2,3	4,8 3,2 2,2	4,7 3,1 2,2	4,5 3,0 2,1	4,4 2,9 2,1	4,3 2,8 2,1	4,2 2,8 2,0	4,1 2,7 2,0	4,0 2,7 2,0	3,9 2,6 1,9	3,8 2,6 1,9	3,7 2,6 1,9	3,9 2,9 2,1	18
19	4,8 3,2 2,3	4,7 3,1 2,2	4,5 3,0 2,1	4,4 2,9 2,1	4,2 2,9 2,1	4,1 2,8 2,0	4,0 2,7 2,0	3,9 2,6 2,0	3,8 2,6 1,9	3,7 2,5 1,9	3,6 2,5 1,9	3,5 2,5 1,9	3,9 2,9 2,1	19
20	4,7 3,1 2,2	4,5 3,0 2,2	4,4 2,9 2,1	4,2 2,9 2,1	4,1 2,8 2,0	4,0 2,7 2,0	3,9 2,6 2,0	3,8 2,6 1,9	3,7 2,5 1,9	3,6 2,5 1,9	3,5 2,4 1,8	3,4 2,4 1,8	3,9 2,8 2,1	20
21	4,5 3,1 2,2	4,4 3,0 2,1	4,2 2,9 2,1	4,0 2,8 2,0	3,9 2,7 2,0	3,8 2,6 2,0	3,7 2,6 1,9	3,6 2,5 1,9	3,6 2,5 1,9	3,5 2,4 1,8	3,4 2,4 1,8	3,3 2,4 1,8	3,8 2,8 2,1	21
22	4,4 3,0 2,2	4,3 2,9 2,1	4,1 2,8 2,1	3,9 2,7 2,0	3,8 2,7 2,0	3,7 2,6 1,9	3,6 2,5 1,9	3,6 2,5 1,8	3,5 2,4 1,8	3,4 2,4 1,8	3,3 2,3 1,8	3,2 2,3 1,8	3,8 2,8 2,1	22
23	4,3 3,0 2,1	4,2 2,9 2,1	4,0 2,8 2,0	3,8 2,7 2,0	3,7 2,6 2,0	3,6 2,5 1,9	3,5 2,5 1,9	3,5 2,4 1,8	3,4 2,4 1,8	3,3 2,3 1,8	3,2 2,3 1,8	3,1 2,3 1,8	3,8 2,8 2,1	23
ν_1 / ν_2	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_2 \ ν_1

Продолжение

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	11	12	ν_1 / ν_2
24	14,0 7,8 4,3	9,3 5,6 3,4	7,6 4,7 3,0	6,6 4,2 2,8	6,0 3,9 2,6	5,6 3,7 2,5	5,3 3,5 2,4	5,0 3,4 2,4	4,9 3,2 2,3	4,7 3,2 2,3	4,6 3,1 2,2	4,4 3,0 2,2	24
25	13,9 7,8 4,2	9,2 5,6 3,4	7,5 4,7 3,0	6,5 4,2 2,8	5,9 3,9 2,6	5,5 3,6 2,5	5,2 3,5 2,4	4,9 3,3 2,3	4,8 3,2 2,3	4,6 3,1 2,2	4,5 3,0 2,2	4,3 3,0 2,2	25
26	13,7 7,7 4,2	9,1 5,5 3,4	7,4 4,6 3,0	6,4 4,1 2,7	5,8 3,8 2,6	5,4 3,6 2,5	5,1 3,4 2,4	4,8 3,3 2,3	4,7 3,2 2,3	4,5 3,1 2,2	4,4 3,0 2,2	4,2 3,0 2,1	26
27	13,6 7,7 4,2	9,0 5,5 3,3	7,3 4,6 3,0	6,3 4,1 2,7	5,7 3,8 2,6	5,3 3,6 2,5	5,1 3,4 2,4	4,8 3,3 2,3	4,7 3,1 2,2	4,5 3,1 2,2	4,4 3,0 2,2	4,2 2,9 2,1	27
28	13,5 7,6 4,2	8,9 5,4 3,3	7,2 4,6 2,9	6,3 4,1 2,7	5,7 3,8 2,6	5,2 3,5 2,4	5,0 3,4 2,4	4,7 3,2 2,3	4,6 3,1 2,2	4,4 3,0 2,2	4,3 2,9 2,1	4,1 2,9 2,1	28
29	13,4 7,6 4,2	8,9 5,4 3,3	7,1 4,5 2,9	6,2 4,0 2,7	5,6 3,7 2,5	5,2 3,5 2,4	5,0 3,3 2,2	4,7 3,2 2,3	4,6 3,1 2,2	4,4 3,0 2,2	4,3 2,9 2,1	4,1 2,9 2,1	29
30	13,3 7,6 4,2	8,8 5,4 3,3	7,1 4,5 2,9	6,1 4,0 2,7	5,5 3,7 2,5	5,1 3,5 2,4	4,9 3,3 2,3	4,6 3,2 2,3	4,5 3,1 2,2	4,3 3,0 2,2	4,2 2,9 2,1	4,0 2,8 2,1	30
ν_2 / ν_1	1	2	3	4	5	6	7	8	9	10	11	12	ν_2 \ ν_1

Продолжение

ν_2 \ ν_1	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_1 / ν_2
24	4,2 2,9 2,1	4,1 2,8 2,1	3,9 2,7 2,0	3,7 2,7 2,0	3,6 2,6 1,9	3,5 2,5 1,9	3,4 2,4 1,9	3,4 2,4 1,8	3,3 2,3 1,8	3,2 2,3 1,8	3,1 2,2 1,7	3,0 2,2 1,7	3,8 3,8 2,1	24
25	4,2 2,9 2,1	4,0 2,8 2,1	3,9 2,7 2,0	3,7 2,7 2,0	3,6 2,5 1,9	3,5 2,4 1,9	3,4 2,4 1,8	3,3 2,3 1,8	3,2 2,3 1,8	3,1 2,3 1,7	3,0 2,2 1,7	2,9 2,2 1,7	3,6 2,8 2,1	25
26	4,1 2,9 2,1	3,9 2,8 2,0	3,8 2,7 2,0	3,6 2,6 1,9	3,5 2,5 1,9	3,4 2,4 1,8	3,3 2,4 1,8	3,2 2,3 1,8	3,1 2,2 1,8	3,0 2,2 1,7	2,9 2,1 1,7	2,8 2,1 1,7	3,6 2,8 2,1	26
27	4,0 2,8 2,1	3,9 2,7 2,0	3,7 2,6 2,0	3,5 2,5 1,9	3,4 2,5 1,9	3,3 2,4 1,8	3,2 2,3 1,8	3,2 2,2 1,8	3,1 2,2 1,7	3,0 2,2 1,7	2,9 2,1 1,7	2,8 2,1 1,7	3,7 2,8 2,1	27
28	4,0 2,8 2,1	3,8 2,7 2,0	3,7 2,6 2,0	3,5 2,5 1,9	3,4 2,4 1,9	3,3 2,3 1,8	3,2 2,3 1,8	3,1 2,2 1,7	3,0 2,2 1,7	2,9 2,1 1,7	2,8 2,1 1,7	2,7 2,1 1,6	3,7 2,8 2,1	28
29	3,9 2,8 2,0	3,8 2,7 2,0	3,6 2,6 1,9	3,4 2,5 1,9	3,3 2,4 1,8	3,2 2,3 1,8	3,1 2,3 1,8	3,0 2,2 1,7	2,9 2,1 1,7	2,8 2,1 1,7	2,7 2,1 1,6	2,6 2,0 1,6	3,7 2,8 2,0	29
30	3,9 2,7 2,0	3,7 2,7 2,0	3,6 2,5 1,9	3,4 2,5 1,9	3,3 2,4 1,8	3,2 2,3 1,8	3,1 2,2 1,8	3,0 2,2 1,7	2,9 2,1 1,7	2,8 2,1 1,7	2,7 2,0 1,6	2,6 2,0 1,6	3,6 2,8 2,0	30
ν_2 / ν_1	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_2 \ ν_1

Продолжение

ν_1 / ν_2	1	2	3	4	5	6	7	8	9	10	11	12	ν_1 / ν_2
32	13,2 7,5 4,1	8,7 5,3 3,3	7,0 4,5 2,9	6,0 4,0 2,7	5,4 3,7 2,5	5,0 3,4 2,4	4,8 3,2 2,3	4,5 3,1 2,2	4,4 3,0 2,2	4,2 2,9 2,1	4,1 2,9 2,1	3,9 2,8 2,1	32
34	13,1 7,4 4,1	8,6 5,3 2,3	7,0 4,4 2,9	6,0 3,9 2,6	5,4 3,6 2,5	5,0 3,4 2,4	4,8 3,2 2,3	4,5 3,1 2,2	4,4 3,0 2,1	4,2 2,9 2,1	4,1 2,8 2,1	3,9 2,8 2,0	34
36	13,0 7,4 4,1	8,6 5,2 3,3	6,9 4,4 2,9	5,9 3,9 2,6	5,3 3,6 2,5	4,9 3,3 2,4	4,7 3,2 2,3	4,4 3,0 2,2	4,3 2,9 2,1	4,1 2,9 2,1	4,0 2,8 2,1	3,8 2,7 2,0	36
38	12,9 7,3 4,1	8,5 5,2 3,2	6,8 4,3 2,8	5,8 3,9 2,6	5,3 3,5 2,5	4,9 3,3 2,3	4,7 3,1 2,3	4,4 3,0 2,2	4,3 2,9 2,1	4,1 2,8 2,1	4,0 2,7 2,1	3,8 2,7 2,0	38
40	12,8 7,3 4,1	8,4 5,2 3,2	5,8 4,3 2,8	5,2 3,8 2,6	4,8 3,5 2,4	4,6 3,3 2,3	4,3 3,1 2,2	4,3 3,0 2,2	4,2 2,9 2,1	4,0 2,8 2,1	3,9 2,7 2,0	3,7 2,7 2,0	40
42	12,7 7,3 4,1	8,3 5,1 3,2	6,7 4,3 2,8	5,7 3,8 2,6	5,2 3,5 2,4	4,8 3,3 2,3	4,6 3,1 2,2	4,3 3,0 2,2	4,2 2,9 2,1	4,0 2,8 2,1	3,9 2,7 2,0	3,7 2,6 2,0	42
44	12,5 7,2 4,1	8,2 5,1 3,2	6,6 4,3 2,8	5,6 3,8 2,6	5,1 3,5 2,4	4,7 3,2 2,3	4,5 3,1 2,2	4,2 2,9 2,2	4,1 2,8 2,1	3,9 2,7 2,0	3,8 2,7 2,0	3,6 2,6 2,0	44
ν_2 / ν_1	1	2	3	4	5	6	7	8	9	10	11	12	ν_2 / ν_1

Продолжение

ν_2 \ ν_1	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_1 / ν_2
32	3,8 2,7 2,0	3,7 2,6 2,0	3,5 2,5 1,9	3,4 2,4 1,9	3,2 2,3 1,8	3,2 2,2 1,8	3,1 2,2 1,7	3,0 2,1 1,7	2,8 2,1 1,7	2,7 2,0 1,6	2,6 2,0 1,6	2,5 2,0 1,6	3,6 2,7 2,0	32
34	3,8 2,7 2,0	3,6 2,6 1,9	3,5 2,5 1,9	3,3 2,4 1,8	3,2 2,3 1,8	3,1 2,2 1,7	3,0 2,1 1,7	2,9 2,1 1,6	2,8 2,0 1,6	2,6 2,0 1,6	2,6 1,9 1,6	2,5 1,9 1,5	3,6 2,7 2,0	34
36	3,7 2,6 2,0	3,6 2,5 1,9	3,4 2,4 1,9	3,3 2,3 1,8	3,1 2,3 1,8	3,1 2,2 1,7	3,0 2,1 1,7	2,9 2,0 1,6	2,7 2,0 1,6	2,6 1,9 1,6	2,5 1,9 1,6	2,4 1,9 1,5	3,6 2,7 2,0	36
38	3,7 2,6 2,0	3,5 2,5 1,9	3,4 2,4 1,9	3,2 2,3 1,8	3,1 2,2 1,8	3,0 2,1 1,7	2,9 2,1 1,7	2,8 2,0 1,6	2,7 2,0 1,6	2,6 1,9 1,6	2,5 1,9 1,5	2,4 1,8 1,5	3,6 2,7 2,0	38
40	3,6 2,6 1,9	3,5 2,5 1,9	3,3 2,4 1,8	3,2 2,3 1,8	3,0 2,2 1,7	3,0 2,1 1,7	2,9 2,0 1,7	2,8 2,0 1,6	2,6 1,9 1,6	2,5 1,9 1,5	2,4 1,8 1,5	2,3 1,8 1,5	3,6 2,7 2,0	40
42	3,6 2,5 1,9	3,4 2,5 1,9	3,3 2,3 1,8	3,1 2,3 1,8	3,0 2,2 1,7	2,9 2,1 1,7	2,8 2,0 1,6	2,7 1,9 1,6	2,6 1,9 1,6	2,4 1,8 1,5	2,4 1,8 1,5	2,3 1,8 1,5	3,6 2,7 2,0	42
44	3,5 2,5 1,9	3,4 2,4 1,9	3,2 2,3 1,8	3,1 2,2 1,8	2,9 2,1 1,7	2,9 2,1 1,7	2,8 2,0 1,6	2,7 1,9 1,6	2,5 1,9 1,6	2,4 1,8 1,5	2,3 1,8 1,5	2,2 1,7 1,5	3,5 2,7 2,0	44
ν_2 / ν_1	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_2 \ ν_1

Продолжение

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	11	12	ν_2 \ ν_1
46	12,4 7,2 4,0	8,1 5,1 3,2	6,5 4,2 2,8	5,6 3,8 2,6	5,0 3,4 2,4	4,6 3,2 2,3	4,4 3,0 2,2	4,1 2,9 2,1	4,0 2,8 2,1	3,8 2,7 2,0	3,7 2,7 2,0	3,5 2,6 2,0	46
48	12,3 7,2 4,0	8,1 5,1 3,2	6,4 4,2 2,8	5,5 3,7 2,6	5,0 3,4 2,4	4,6 3,2 2,3	4,4 3,0 2,2	4,1 2,9 2,1	4,0 2,8 2,1	3,8 2,7 2,0	3,7 2,6 2,0	3,5 2,6 2,0	48
50	12,2 7,2 4,0	8,0 5,1 3,2	6,4 4,2 2,8	5,4 3,7 2,6	4,9 3,4 2,4	4,5 3,2 2,3	4,3 3,0 2,2	4,0 2,9 2,1	3,9 2,8 2,1	3,7 2,7 2,0	3,6 2,6 2,0	3,4 2,6 1,9	52
55	12,1 7,1 4,0	7,9 5,0 3,2	6,3 4,1 2,8	5,4 3,7 2,5	4,9 3,4 2,4	4,5 3,1 2,3	4,3 3,0 2,2	4,0 2,8 2,1	3,9 2,7 2,0	3,7 2,7 2,0	3,6 2,6 2,0	3,4 2,5 1,9	55
60	12,0 7,1 4,0	7,8 5,0 3,1	6,2 4,1 2,8	5,3 3,6 2,5	4,8 3,3 2,4	4,4 3,1 2,2	4,2 2,9 2,2	3,9 2,8 2,1	3,8 2,7 2,0	3,6 2,6 2,0	3,5 2,6 1,9	3,3 2,5 1,9	60
65	11,9 7,0 4,0	7,7 5,0 3,1	6,1 4,1 2,7	5,2 3,6 2,5	4,7 3,3 2,4	4,3 3,1 2,2	4,1 2,9 2,1	3,8 2,8 2,1	3,7 2,7 2,0	3,5 2,6 2,0	3,4 2,5 1,9	3,2 2,5 1,9	65
70	11,6 7,0 4,0	7,6 4,9 3,1	6,0 4,1 2,7	5,2 3,6 2,5	4,7 3,3 2,3	4,3 3,1 2,2	4,1 2,9 2,1	3,8 2,8 2,1	3,7 2,7 2,0	3,5 2,6 2,0	3,4 2,5 1,9	3,2 2,4 1,9	70
ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	11	12	ν_2 \ ν_1

Продолжение

ν_1 / ν_2	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_1 / ν_2
46	3,4 2,5 1,9	3,3 2,4 1,9	3,1 2,3 1,8	3,0 2,2 1,7	2,8 2,1 1,7	2,8 2,0 1,6	2,7 2,0 1,6	2,6 1,9 1,6	2,4 1,9 1,5	2,3 1,8 1,5	2,2 1,8 1,5	2,1 1,7 1,5	3,5 2,7 2,0	46
48	3,4 2,5 1,9	3,3 2,4 1,9	3,1 2,3 1,8	3,0 2,2 1,7	2,8 2,1 1,7	2,8 2,0 1,6	2,7 2,0 1,6	2,6 1,9 1,6	2,4 1,8 1,5	2,3 1,8 1,5	2,2 1,7 1,5	2,1 1,7 1,4	3,5 2,7 2,0	48
50	3,3 2,5 1,9	3,2 2,4 1,8	3,0 2,3 1,8	2,9 2,2 1,7	2,7 2,1 1,7	2,7 2,0 1,6	2,6 1,9 1,6	2,5 1,9 1,5	2,3 1,8 1,5	2,2 1,8 1,5	2,1 1,7 1,5	2,0 1,7 1,4	3,5 2,7 2,0	50
55	3,3 2,5 1,9	3,2 2,3 1,8	3,0 2,2 1,8	2,9 2,1 1,7	2,7 2,1 1,7	2,7 2,0 1,6	2,6 1,9 1,6	2,5 1,8 1,5	2,3 1,8 1,5	2,2 1,7 1,5	2,1 1,7 1,4	2,0 1,6 1,4	3,5 2,7 2,0	55
60	3,2 2,4 1,9	3,1 2,3 1,8	2,9 2,2 1,7	2,8 2,1 1,7	2,6 2,0 1,6	2,6 1,9 1,6	2,5 1,9 1,6	2,4 1,8 1,5	2,2 1,7 1,5	2,1 1,7 1,4	2,0 1,6 1,4	1,9 1,6 1,4	2,4 2,7 2,0	60
65	3,1 2,4 1,8	3,0 2,3 1,8	2,8 2,2 1,7	2,7 2,1 1,7	2,5 2,0 1,6	2,5 1,9 1,6	2,4 1,8 1,5	2,3 1,8 1,5	2,1 1,7 1,5	2,0 1,6 1,4	1,9 1,6 1,4	1,8 1,6 1,4	3,4 2,6 2,0	65
70	3,1 2,3 1,8	3,0 2,3 1,8	2,8 2,1 1,7	2,7 2,1 1,7	2,5 2,0 1,6	2,5 1,9 1,6	2,3 1,8 1,5	2,2 1,7 1,5	2,1 1,7 1,4	2,0 1,6 1,4	1,8 1,6 1,4	1,7 1,5 1,3	3,4 2,6 2,0	70
ν_2 / ν_1	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_2 / ν_1

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	11	12	ν_1 / ν_2
80	11,6 7,0 4,0	7,5 4,9 3,1	6,0 4,0 2,7	5,1 3,6 2,5	4,6 3,2 2,3	4,2 3,0 2,2	5,0 2,9 2,1	3,7 2,7 2,0	3,6 2,6 2,0	3,4 2,5 1,9	3,3 2,5 1,9	3,1 2,4 1,9	80
100	11,5 6,9 3,9	7,4 4,8 3,1	5,9 4,0 2,7	5,0 3,5 2,5	4,5 3,2 2,3	4,1 3,0 2,2	3,9 2,8 2,1	3,7 2,7 2,0	3,6 2,6 2,0	3,4 2,5 1,9	3,3 2,4 1,9	3,1 2,4 1,8	100
125	11,4 6,8 3,9	7,4 4,8 3,1	5,8 3,9 2,7	5,0 3,5 2,4	4,5 3,2 2,3	4,1 2,9 2,2	3,9 2,8 2,1	3,6 2,6 2,0	3,5 2,6 1,9	3,3 2,5 1,9	3,2 2,4 1,9	3,0 2,3 1,8	125
150	11,3 6,8 3,9	7,3 4,7 3,1	5,7 3,9 2,7	4,9 3,4 2,4	4,4 3,1 2,3	4,0 2,9 2,2	3,8 2,8 2,1	3,5 2,6 2,0	3,4 2,5 1,9	3,2 2,4 1,9	3,1 2,4 1,8	2,9 2,3 1,8	150
200	11,2 6,8 3,9	7,2 4,7 3,0	5,6 3,9 2,6	4,8 3,4 2,4	4,3 3,2 2,3	3,9 2,9 2,1	3,7 2,7 2,0	3,5 2,6 2,0	3,4 2,5 1,9	3,2 2,4 1,9	3,1 2,3 1,8	2,9 2,3 1,8	200
400	11,0 6,7 3,9	7,1 4,7 3,0	5,6 3,8 2,6	4,7 3,4 2,4	4,2 3,1 2,2	3,8 2,8 2,1	3,6 2,7 2,0	3,4 2,5 2,0	3,3 2,5 1,9	3,1 2,4 1,8	3,0 2,3 1,8	2,8 2,2 1,8	400
1000	10,9 6,7 3,8	7,0 4,6 3,0	5,5 3,8 2,6	4,7 3,4 2,4	4,2 3,3 2,2	3,8 3,0 2,1	3,6 2,8 2,0	3,4 2,7 1,9	3,3 2,5 1,9	3,1 2,4 1,8	3,0 2,3 1,8	2,8 2,3 1,8	1000
	10,8 6,6 3,8	6,9 4,6 3,0	5,4 3,8 2,6	4,6 3,3 2,4	4,1 3,0 2,2	3,7 2,8 2,1	3,5 2,6 2,0	3,3 2,5 1,9	3,2 2,4 1,9	3,0 2,3 1,8	2,9 2,2 1,8	2,7 2,2 1,7	
ν_2 / ν_1	1	2	3	4	5	6	7	8	9	10	11	12	ν_2 \ ν_1

ν_1 / ν_2	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_1 / ν_2
80	3,0 2,3 1,8	2,9 2,2 1,8	2,7 2,1 1,7	2,6 2,0 1,6	2,4 1,9 1,6	2,4 1,8 1,5	2,3 1,8 1,5	2,2 1,7 1,4	2,0 1,6 1,4	1,9 1,6 1,4	1,8 1,5 1,3	1,7 1,5 1,3	3,4 2,6 2,0	80
100	3,0 2,3 1,8	2,8 2,2 1,7	2,7 2,1 1,7	2,5 2,0 1,6	2,4 1,9 1,6	2,3 1,8 1,5	2,2 1,7 1,5	2,1 1,6 1,4	1,9 1,6 1,4	1,8 1,5 1,3	1,7 1,5 1,3	1,6 1,4 1,3	3,4 2,6 2,0	100
125	2,9 2,2 1,8	2,8 2,1 1,7	2,6 2,0 1,6	2,5 1,9 1,6	2,3 1,8 1,5	2,3 1,7 1,5	2,1 1,7 1,4	2,0 1,6 1,4	1,9 1,5 1,4	1,8 1,5 1,3	1,6 1,4 1,3	1,5 1,4 1,2	3,4 2,6 2,0	125
150	2,8 2,2 1,8	2,7 2,1 1,7	2,5 2,0 1,6	2,4 1,9 1,6	2,2 1,8 1,5	2,2 1,7 1,5	2,0 1,7 1,4	1,9 1,6 1,4	1,8 1,5 1,3	1,7 1,4 1,3	1,5 1,4 1,2	1,4 1,3 1,2	3,4 2,6 2,0	150
200	2,8 2,2 1,7	2,6 2,1 1,7	2,5 2,0 1,6	2,3 1,9 1,6	2,2 1,8 1,5	2,1 1,7 1,4	1,9 1,6 1,4	1,8 1,5 1,3	1,7 1,5 1,3	1,6 1,4 1,3	1,4 1,3 1,2	1,3 1,3 1,2	3,3 2,6 2,0	200
400	2,7 2,1 1,7	2,5 2,0 1,7	2,4 1,9 1,6	2,2 1,8 1,5	2,1 1,7 1,5	2,0 1,6 1,4	1,9 1,6 1,4	1,8 1,5 1,3	1,6 1,4 1,3	1,5 1,3 1,2	1,4 1,2 1,2	1,2 1,3 1,1	3,3 2,6 2,0	400
1000	2,7 2,2 1,7	2,5 2,1 1,6	2,4 2,0 1,6	2,2 1,9 1,5	2,1 1,8 1,5	2,0 1,7 1,4	1,8 1,6 1,4	1,7 1,5 1,3	1,6 1,4 1,3	1,5 1,4 1,2	1,3 1,2 1,1	1,2 1,1 1,1	3,3 2,6 2,0	1000
	2,6 2,1 1,7	2,4 2,0 1,6	2,3 1,9 1,6	2,1 1,8 1,5	2,0 1,7 1,5	1,9 1,6 1,4	1,7 1,5 1,3	1,6 1,4 1,3	1,5 1,4 1,2	1,4 1,2 1,2	1,2 1,1 1,1	1,1 1,0 1,0	3,3 2,6 2,0	
ν_2 / ν_1	14	16	20	24	30	40	50	75	100	200	500	∞	t	ν_2 / ν_1

ОГЛАВЛЕНИЕ

Стр.

Предисловие 3
Глава I. Предмет и метод вариационной статистики. Место математического метода в биологических исследованиях . . 5
Глава II. Основные статистические понятия и задачи вариационной статистики 9
Глава III. Средние величины и их свойства. Моменты и их использование при вычислениях 21
1. Средняя арифметическая M или $\bar{x}$ 21
2. Взвешенная средняя арифметическая $\bar{x}_{взв}$ 23
3. Средняя геометрическая G 25
4. Средняя квадратическая S 26
5. Мода Mo 28
6. Медиана или срединное значение вариационного ряда Me . 29
7. Моменты статистических величин 32
8. Техника обработки вариационных рядов 33
Глава IV. Изменчивость и методы ее измерения 37
1. Определение степени изменчивости методом лимитов и средним квадратическим отклонением σ 38
2. Поправка Шеппарда на внутриклассовые колебания . . 41
3. Вычисление σ методом сумм 41
4. Вычисление σ для малых выборок при однозначном или двузначном значениях вариантов 43
5. Вычисление σ для малых выборок при дробных или многозначных значениях вариантов 44
6. Способ получения суммарной средней арифметической M и среднего квадратического отклонения σ 46
7. Вычисление среднего квадратического отклонения σ для альтернативных признаков 47
8. Измерение вариации или изменчивости с помощью коэффициента вариации C (или V) 49
Глава V. Графический метод анализа вариационных рядов и типы распределений (нормальное, биноминальное, Пуассона, Шарлье и трансгрессии) 51
1. Построение эмпирической вариационной кривой при прерывистом значении признака 51
2. Построение кривой с непрерывным значением признака . 51
3. Нормальное распределение и нормальная кривая . . . 53
4. Биноминальное распределение 60
5. Распределение Пуассона 61
6. Асимметричные вариационные кривые и распределение Шарлье. Мера асимметрии As 61

7. Эксцессивные вариационные кривые. Мера эксцесса *Ex* . 63
8. Двухвершинные и многовершинные кривые 65
9. Трансгрессивная изменчивость 66
10. Метод комбинированных признаков 67
Глава VI Статистические связи и методы их вычисления 70
1. Типы связей и общие замечания о связях 70
2. Коэффициент корреляции *r* и техника его вычисления при различных условиях 74
а. Вычисление коэффициента корреляции для малой выборки при однозначных и двузначных значениях вариантов 75
б. Вычисление коэффициента корреляции при многозначных или дробных значениях вариантов 77
в Вычисление коэффициента корреляции для альтернативных признаков 79
г. Определение коэффициента корреляции между количественными признаками при большом числе наблюдений 80
3. Регрессия как показатель связи *R* (уравнение регрессии, коэффициент регрессии, линия регрессии) 86
4. Полихорический показатель связи ρ (второй коэффициент сопряженности Чупрова) 91
5. Бисериальный показатель связи *r* 94
6. Криволинейные типы связи 97
а. Корреляционное отношение η 97
б Вычисление корреляционного отношения для малой выборки 98
в. Вычисление корреляционного отношения для большой выборки 105
Глава VII Статистические ошибки (*мера точности статистических характеристик*) 114
1. Типы ошибок и методы их уменьшения 114
2. Формулы ошибок статистических величин и способы их использования 119
3. Метод хи-квадрат (χ^2) как мера вероятности совпадения частот разных вариационных рядов (эмпирического и теоретического) 138
4. Выравнивание данных методом *скользящей средней* . . 145
Глава VIII. Дисперсионный анализ 147
1. Основные положения дисперсионного анализа 147
2. Техника обработки статистического комплекса при дисперсионном анализе 158
А. Равномерные и пропорциональные комплексы . . . 158
Б. Неравномерные статистические *комплексы* 182
В. Дисперсионный анализ качественных признаков в однофакторных, двухфакторных и трехфакторных комплексах 196
Г. Использование дисперсионного анализа для сравнения средних арифметических внутри комплекса и для определения достоверности разницы между ними 203

Евгения Константиновна Меркурьева

ОСНОВЫ БИОМЕТРИИ

Редакторы Сысоева Н В, Малахов Ф Н
Техред Ермаков М. С. Корректор Гаврилюк А. И.

Сдано в набор 16/VI 1962 г Подписано к печати 19/II 1963 г.
Л-57108 Формат 60×90 1/16 Печ л 15,0 Уч-изд л 12,97
Заказ 158 Тираж 1250 экз Цена 60 коп.

Издательство Московского университета, Москва, Ленинские горы
Типография Изд-ва МГУ, Москва, Ленинские горы

Цена 60 коп.

Цена 60 коп.