

МИНИСТЕРСТВО ОБЩЕГО И ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
САМАРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Биологический факультет

**МАТЕМАТИЧЕСКИЕ МЕТОДЫ
В БОТАНИЧЕСКИХ И ЭКОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ**
Учебно-методические указания

Издательство «Самарский университет»
1998

*Печатается по решению Редакционно-издательского совета
Самарского государственного университета*

Учебно-методические указания предназначены непосредственно для студентов 3 курса специальности «Биология», специализации «Экология, ботаника и охраны природы», которые изучают данную дисциплину в рамках спецкурса, а также для студентов других курсов и специализаций, принимающих участие в учебно-исследовательской работе и выполняющих курсовые и дипломные работы.

Данное, второе, издание учебно-методических указаний значительно переработано, расширено за счет дополнительных примеров и иллюстраций и включения новых разделов и приложений.

Учебно-методические указания могут быть также полезны в работе начинающих исследователей, педагогов, школьников-старшеклассников.

Составитель доц., канд.биол.наук. Л.М.Кавеленова
Рецензент доц. Г.Л.Рытов

© Кавеленова Л.М.,
составление, 1998

ПРЕДИСЛОВИЕ

Данное руководство представляет собой второе издание учебно-методических указаний по математическим методам в ботанических и экологических исследованиях. Первое издание, опубликованное в 1994 году, использовалось в преподавании одноименного спецкурса на специализации «Экология, ботаника и охрана природы» и предлагалось в помощь студентам дневного и вечернего отделений при выполнении ими курсовых и дипломных работ. Тираж первого издания оказался полностью исчерпан менее чем за 3 года. Это убедило нас, что несмотря на имеющуюся литературу по биометрии студентам по-прежнему необходимы достаточно простые руководства малого объема, которые позволяют применить аппарат статистики к решению практической исследовательской задачи. Предлагаемые указания содержат описание методов математической обработки результатов ботанических экспериментов и полевых исследований, наиболее распространенных и доступных при выполнении студентами курсовых и дипломных работ – тот весьма скудный минимум, которым сегодня обязательно должен владеть биолог вне зависимости от его специализации.

К сожалению, даже для диссертационных работ, представляемых к защите в области биологии и медицины, уровень использования прикладной статистики не отвечает современным возможностям. Интересный анализ ситуации приводят В.П.Леонов и П.В.Ижевский [1]: так, на основании анализа более 200 диссертационных работ, а также публикаций в 10 наиболее известных биомедицинских журналах за 3 года они отмечают использование в 30-40% работ только критерия Стьюдента, который зачастую неприменим для конкретных данных. Относительная частота встречаемости корреляционного анализа – 7%, в 62% работ используется выражение « $p < \dots$ », в 79% работ данные приводят в форме « $M \pm m$ ». Для 20% работ статистика вообще не применялась! Отметим, что при проведении этого анализа авторы не учитывали обзорных и лекционных статей, а также работ, не использовавших выборочного метода анализа наблюдений. Характерно, что около 15 лет назад аналогичным образом обстояла ситуация и в зарубежной медицинской и биологической науках. Однако с тех пор, в силу ряда причин, уровень использования прикладной статистики за рубежом существенно повысился. Картина, отражающая использование различных групп математических методов отечественными и зарубежными учеными, которые публикуются в журналах «Экология» и «Ecology» соответственно, достаточно неутешительна [1]. Поэтому активное освоение языка статистики студентами-биологами становится тем более необходимым.

Составитель указаний видит свою непосредственную задачу прежде всего в том, чтобы еще раз напомнить читателю хорошо известные из курса биометрии приемы обработки данных, интерпретации полученных результатов в ботанико-экологическом аспекте, а также ознакомить с некоторыми сугубо ботаническими методами использования математического аппарата.

Самые распространенные задачи, для решения которых исследователь-ботаник привлекает математический аппарат, можно охарактеризовать следующим образом:

- определение показателей, характеризующих ряд однородных чисел - полученных данных,
- оценка степени достоверности - значимости - различий между сериями измерений, то есть вариантами опыта,
- установление по двум и более сопряженным рядам показателей связи либо установление формы зависимости одного признака от другого, характера зависимости, измеряемого признака от внешних условий [2].

Существуют и специальные способы применения математических методов в конкретных флористических, геоботанических и пр. исследованиях.

При подготовке указаний использовалось несколько руководств по математическим методам в биологии и ботанике, к которым следует обратиться при необходимости более углубленного изучения сложных приемов использования математического аппарата [2-6]. Мы придерживались изложения материала, наиболее доступного и широко употребительного в практике, в форме конспекта, предполагая знакомство читателя с основами биометрии.

В настоящем издании мы шире использовали привлечение в учебных целях конкретных фактов, приводимых исследователями-ботаниками в публикациях, в качестве иллюстраций. Как правило, эти фрагменты пособия выделяются курсивом. Кроме того, некоторое внимание мы уделили графическим приемам представления материалов (графики, диаграммы).

Издание не претендует на полную освещенность современных математических методов в ботанических и экологических исследованиях, однако надеемся, что оно может быть полезно в практической работе начинающим исследователям - ботаникам, физиологам растений, экологам: студентам, стажерам, аспирантам.

ГЛАВА 1. ВАРИАЦИОННЫЕ РЯДЫ И ИХ СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ

1.1. Типы изменчивости признака:

количественная и качественная изменчивость

Количественная изменчивость - изменчивость, при которой различия между вариантами выражаются количеством – массой растений и их органов, высотой, урожаем и др. Она может быть прерывистой (дискретной) или непрерывной. В первом случае признаки варьируют и отличаются на определенную величину, выражаемую целыми числами. Во втором случае значения признака отличаются друг от друга на любую малую величину в зависимости от степени точности измерения.

К настоящему времени известно, что различные количественные признаки растений в норме обладают неодинаковыми уровнями изменчивости, зависящими от видовых и популяционных особенностей объектов. Так, для древесных растений уровни изменчивости некоторых признаков определяются следующим образом (по С.А.Мамаеву, А.К.Махневу [7]):

Таблица 1

Уровни изменчивости признаков древесных пород (из [7])

(низкий – до 12 %, средний – 13-20%, высокий – 21-40%, очень высокий – более 40%)

Тип признаков	Признак	Уровень изменчивости
Структурные признаки вегетативных органов	Высота ствола	Низкий и средний
	Диаметр ствола	Средний
	Протяженность кроны	Высокий
	Длина листьев	Средний и низкий
Структурные признаки генеративных органов	Размеры клеток	Низкий
	Длина плодов	Низкий и средний
	Вес плодов	Высокий
	Вес семян	Средний
Функциональные признаки	Окраска семян	Высокий
	Размер пыльцевых зерен	Низкий
	Обилие плодоношения	Очень высокий
	Абсолютная всхожесть семян	Средний
Биохимические признаки	Термостойкость семян	Очень высокий
	Термостойкость листьев	Очень высокий
	Энергия проросточного роста	Высокий
	Содержание воды в листьях	Низкий
	Содержание хлорофилла	Средний
	Зольность листьев	Низкий и средний

Качественная, или атрибутивная, изменчивость имеет место при качественной оценке явления, когда одни из объектов наблюдения (частей выборки) могут иметь тот или иной показатель, а другие его не имеют. Например, такими признаками будут определенная окраска частей околоцветника, морфологические особенности органов растения, повреждение патогенами и др. Частным случаем для качественной выступает **альтернативная изменчивость**, подразумевающая наличие либо отсутствие рассматриваемого признака как одной из двух взаимоисключающих возможностей. Здесь до-

статочно вспомнить условия обыкновенных задач по генетике, в которых может встречаться как альтернативная изменчивость (при полном доминировании – желтые либо зеленые, гладкие либо морщинистые горошины в опытах Грегора Менделя), так и более сложные варианты (например, при неполном доминировании – красная, розовая или белая окраска околоцветника у цветка ночной красавицы).

Далее мы рассмотрим конкретные случаи использования приемов математической обработки данных в случае количественной и качественной изменчивости.

1.2. Генеральная и выборочная совокупность.

Вариационный ряд и его построение

Генеральная совокупность в общем плане – это группа подлежащих изучению объектов определенной категории, которая, как правило, достаточно велика и теоретически имеет объем, равный бесконечности. Исследователь работает с ее частью – *выборочной совокупностью*, которая характеризуется следующими чертами: формируется в случайном порядке для достижения условия репрезентативности, исследуется для характеристики выборочной и генеральной совокупностей.

В процессе наблюдения получают сведения о численной величине изучаемого признака для каждого члена выборочной совокупности. При первичной обработке данных располагают все значения признака варианты в порядке возрастания или убывания – *ранжируют ряд вариант*. При этом обнаруживается, что каждое значение встречается разное число раз. Повторяемость значений характеризуют частоты признака f , сумма всех частот $\sum f$ равна объему выборки, то есть числу членов ряда N . В результате такой обработки получают *вариационный ряд*, или *ряд распределения*. Для его построения необходимо:

- 1) найти для выборки минимальную и максимальную варианты;
- 2) вычислить разность между ними – *размах варьирования* R ;
- 3) определить число групп (классов) K , на которые следует разбить вариационный ряд. В простейшем случае $K = \sqrt{N}$, можно также использовать формулу Штургеса: $K = 1 + 3.32 * \lg(N)$, в обоих случаях N – объем выборки;
- 4) вычислить групповой (классовый) интервал:

$$L = \frac{X_{\max} - X_{\min}}{K},$$

где $X_{\max}$ и $X_{\min}$ соответственно – минимальное и максимальное значения признака;

- 5) определить начало каждого группового интервала (путем прибавления к минимальному значению признака величины группового интервала – для первого класса, двух величин – для второго и т.д.). Конец соответствующего группового интервала должен быть меньше начала следующего интервала на величину, равную точности измерения признака;
- 6) разнести все варианты выборки последовательно по классам.

1.3. Гистограмма и ее использование

Графически вариационный ряд представляют в виде ступенчатой кривой - *гистограммы*, которая показывает распределение признака в данной совокупности (рис. 1).

Графическое построение гистограммы заключается в том, что по оси абсцисс откладываются границы интервалов, на каждом из интервалов как на основании строится прямоугольник, высота которого равна абсолютной или относительной частоте каждого интервала.

Построение гистограмм и их анализ - один из наиболее простых и доступных методов работы с экспериментальными данными. По виду гистограммы можно получить первичную информацию об основных тенденциях изучаемого процесса или объекта.

Так, смещение центра распределения и несимметричность гистограммы говорят об определенной нестабильности, влиянии внешних факторов (рис. 1B). Плосковершинность распределения - один из показателей неустойчивости исследуемого процесса (рис. 1C), а наличие двух выраженных вершин на гистограмме может соответствовать ситуации, когда в одной выборке оказались случайно объединены принципиально разные по какому-то признаку объекты (рис. 1D). Несмотря на кажущуюся примитивность этих заключений, они тем не менее могут быть использованы в практике обсуждения ваших экспериментальных данных. Ниже мы рассмотрим примеры конкретного применения гистограмм в ботанических исследованиях.



Рис.1. Гистограмма – одна из форм графического изображения вариационного ряда (см. пояснения в тексте).

1.3.1. Гистограмма возрастного спектра ценопопуляции

Ценопопуляция - популяция, находящаяся в пределах биоценоза как устойчивой экосистемы. Ее возрастной спектр представляет собой распределение особей ценопопуляции по возрастам и фазам развития. У растений выделяют особи: *латентные* - L (покоящиеся - семена, луковицы, корневища, споры), *инфантные* - I (всходы), *ювенильные* - J (юношеские, интенсивно растущие), *имматурные* - Im (переход к взрослому состоянию), *виргинильные* - V (хорошо развитые, но еще не цветущие и не плодоносящие), *генерирующие* - G

(спороносящие и плодоносящие) и сенильные - S (старческие, со сниженной интенсивностью фотосинтеза, слабо или совсем не плодоносящие) [8]. Произведя подсчет особей популяции с одновременным отнесением их к соответствующим группам, получают данные для построения возрастного спектра ценопопуляции, оформляют его в форме гистограммы, по виду которой можно сделать выводы о состоянии ценопопуляции в данном сообществе. Так, преобладание особей «молодых» стадий может соответствовать инвазионной популяции, только внедряющейся в данное сообщество, преобладание «старых» при отсутствии всходов и ювенильных особей - о том, что не происходит возобновления данной популяции.

В качестве иллюстрации того, как применяется анализ гистограмм - возрастных спектров ценопопуляции, рассмотрим результаты, опубликованные в статье П.Л.Горчаковского, В.Н.Зуевой [9]. Они изучали четыре эндемичных скально-горностепных вида, принадлежащих к роду *Astragalus*. При анализе динамики возрастной структуры популяций для астрагала карелинского получены следующие данные (рис.2).

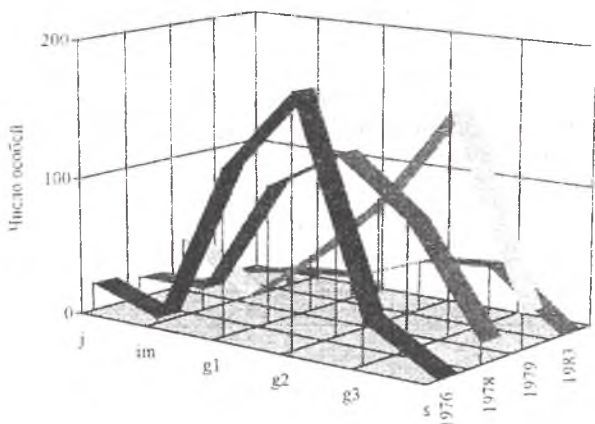


Рис. 2 Динамика возрастной структуры популяции астрагала карелинского (построено по данным [9]).

За восемь лет до начала наблюдений на сопке была проведена посадка сосны, что сопровождалось снятием растительного покрова. На свободных участках началось формирование популяции астрагала.

К началу наблюдений в 1976 г. она была неполноценной (см. рис.2), в 1978г. характеризовалась как средневозрастная полноценная, так как появились сенильные особи. В 1979 г. популяция перешла в разряд старых и вновь стала неполноценной из-за отсутствия имматурных особей.

К 1983 г. в данной популяции отсутствовали уже и ювенильные особи, то есть старение популяции продолжалось. Отмечалось также постепенное снижение плотности популяции. Авторы сделали вывод о том, что в связи с

увеличением сомкнутости травяного покрова и возрастающей конкуренции со стороны степных дерновинных злаков произошло угасание популяции астрагала.

1.3.2. Оценка жизненного состояния древесных растений по В.А. Алексееву и анализ гистограммы состояния популяции

В.А.Алексеев [10] предложил следующую шкалу категорий жизненного состояния деревьев по визуальным характеристикам кроны:

1. Здоровое дерево. Не имеет внешних повреждений кроны и ствола. густота кроны обычная для господствующих деревьев, мертвые и отмирающие ветви сосредоточены в нижней части кроны и отсутствуют в верхней ее половине. Закончившие рост листья и хвоя зеленого или темно-зеленого цвета, их продолжительность жизни типична для региона. Повреждения листьев и хвой незначительны (<10%) и не сказываются на состоянии дерева.

2. Поврежденное (ослабленное) дерево. Обязателен хотя бы один из следующих признаков: снижение густоты кроны на 30% за счет преждевременного опадения или недоразвития листьев (хвой) или изреживания скелетной части кроны; наличие 30% мертвых и (или) усыхающих ветвей в верхней половине кроны; повреждение (объедание, ожог, хлорозы, некрозы и т.д.) и выключение из ассимиляционной деятельности 30% листовой поверхности.

3. Сильно поврежденное (сильно ослабленное) дерево. Обязателен хотя бы один из следующих признаков: снижение густоты облиствения кроны на 60% за счет преждевременного опадения листьев (хвой) или изреживания скелетной части кроны; наличие 60% мертвых и (или) усыхающих ветвей в верхней половине кроны; повреждение различными факторами и выключение из ассимилирующей деятельности 60% площади листьев; отмирание верхушки кроны

4. Отмирающее дерево. Крона разрушена, ее густота - менее 15-20% по сравнению со здоровой; >70% ветвей, в том числе в верхней половине, сухие или бледно-зеленого, желтоватого, оранжево-красного цвета. Некрозы белесого, коричневого или черного цвета. При загрязнении атмосферы большая часть некротизированных листьев быстро отмирает. В комлевой и средней части ствола возможны признаки заселения стволовыми вредителями.

5а. Свежий сухостой. Деревья, погибшие менее года назад. У них возможны остатки сухой хвой или листьев, кора и мелкие ветви часто бывают целы. Как правило, заселены насекомыми-ксилофагами.

5б. Старый сухостой. Деревья, погибшие в прошлые годы. Постепенно утрачивают ветви и кору.

В результате пересчета и оценки деревьев данного насаждения, как и при анализе возрастного спектра ценопопуляции, мы будем располагать рядом данных, представляющих собой подобие ранжированного ряда. Возможно построение гистограммы и ее анализ. При преобладании в выборке деревьев классов 3, 4, 5а можно сделать вывод о недостаточной устойчивости данного вида в конкретных условиях или нецелесообразности его использования в озеленении. («Двувершинность» (преобладание классов 1 и 3 при практическом отсутствии класса, например) может объясняться неоднородностью выборки, на которую не обратил внимания наблюдатель, и т.д.

Автор предложил также рассчитывать индексы состояния древостоев по числу деревьев, относящихся к различным классам:

$$L = \frac{100n_1 + 70n_2 + 40n_3 + 5n_4}{N},$$

где L - относительное жизненное состояние древостоя, n_1 - число здоровых деревьев, n_2 - ослабленных, n_3 - сильно ослабленных, n_4 - отмирающих деревьев, N - общее число деревьев, включая сухостой, на пробной площади.

Приведем в качестве иллюстрации результаты, полученные нами совместно с Н.В. Прохоровой при обследовании состояния деревьев ели колючей, произрастающих на территории завода Металлист (г. Самара, 1989г.). Поскольку озеленение промплощадки проводили постоянно в течение последних трех десятилетий, то к моменту наблюдений в насаждениях присутствовали деревья, принадлежащие к двум группам возраста: молодые (до 10 лет) и взрослые (более 20 лет). Относя их к определенным классам по методике Алексеева, мы обнаружили различные спектры состояния (рис. 3). Контрольные группы растений обоих возрастов, произрастающие вне промплощадки, характеризовались лучшим состоянием.

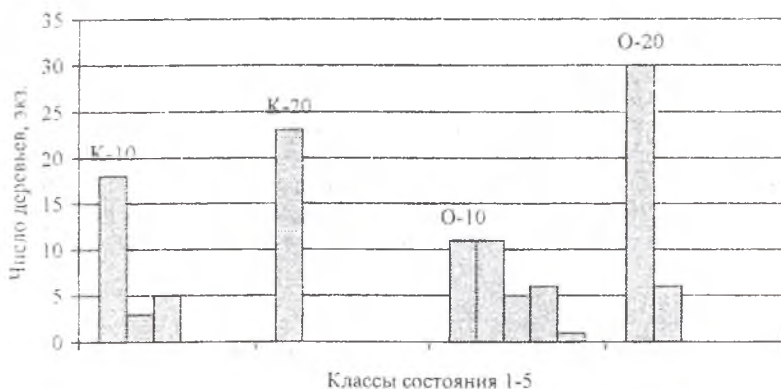


Рис. 3. Классы возрастного состояния деревьев ели колючей на промплощадке завода "Металлист" (О) и в парковой зоне (К).

Можно также отметить, что молодые растения ели являются более уязвимыми в условиях техногенного загрязнения, о чем свидетельствует наличие классов состояния 3-5 в составе спектров. Изменения спектра с достижением 20-летнего возраста мы объясняем тем, что, пережив некий критический период, ослабленные растения погибают и выбраковываются из насаждения.

1.4. Выбровка «выскакивающих» вариант

Наибольшее значение признака в выборке представляет собой максимальную варианту, а наименьшее - минимальную. Если в ранжированном

рядом значение максимальной или минимальной варианты чрезмерно отличается от соседних чисел ряда, необходимо проверить их принадлежность к данной совокупности. чтобы ответить на вопрос - следует ли исключить эти варианты из выборки, считая их результатом ошибки счета или наблюдения.

Для *максимальной* варианты при нормальном или близком к нему распределении вариант применяется формула:

$$V_N = \frac{X_N - X_{N-1}}{X_N - X_2},$$

где V_N - критерий принадлежности максимальной варианты к совокупности, X_N - максимальная варианта, X_{N-1} - варианта, следующая по величине за максимальной, X_2 - варианта, стоящая в ранжированном ряду рядом с минимальной.

Для *минимальной* варианты при распределении, не сильно отличающемся от нормального, применяется формула:

$$V_1 = \frac{X_2 - X_1}{X_{N-1} - X_1},$$

где X_1 - минимальная варианта. Найденное по формулам значение критерия V_1 сравнивают с табличным для данного числа вариантов, если полученное значение меньше табличного - отбрасывать варианту не следует (см. табл. 1 Приложения).

Например, при оценке аллелопатической активности вытяжек из листового опада, биотестировании влияния химических соединений на всхожесть семян и рост корней проростков вы обнаружили, что в обширной выборке с преобладанием нормальных проростков с корнями более 30мм имеется 2-3 проклюнувшихся семени, у которых длина корня не превышает 1мм. Как правильно поступить при расчете средней длины корня по данным опыта - учитывать все проростки или отбросить 2-3, которые вас не устраивают? Примените математическую формулу, вам не придется гадать.

1.5. Статистические характеристики распределения случайной величины

Статистические показатели, используемые для оценки данных в выборке, можно отнести к трем группам величин: *мерам положения* или средним величинам (средняя арифметическая, средняя геометрическая и т.д., медиана, мода), *мерам рассеяния* или показателям изменчивости (лимиты, дисперсия, среднее квадратическое отклонение и коэффициент вариации), *мерам формы* (коэффициенты асимметрии и эксцесса).

1.5.1. Меры положения: средняя арифметическая, медиана, мода, средняя квадратическая, средняя кубическая, средняя гармоническая

Средняя арифметическая - основная характеристика выборки, она отражает уровень, по отношению к которому колеблются значения вариант. Если сумму всех вариант ряда ($X_1 + X_2 + X_3 + \dots + X_N$) мы обозначим как ΣX , то при числе вариантов, равном N , средняя арифметическая равна:

$$\bar{X} = \Sigma X / N.$$

Пожалуй, средняя арифметическая – наиболее известная из характеристик выборки. Тем не менее уместно сделать два замечания. Во-первых, она не должна выступать в качестве единственного показателя. Для средней арифметической необходимо, например, также рассчитать ошибку и оценить достоверность, о чем расскажем позднее (см .1.5.2). Во-вторых, средняя арифметическая не является универсальным показателем. Существует несколько близких «средних», которые вычисляют для конкретного типа данных. Рассмотрим их применение.

При выведении средних величин для значений площади и объема необходимо согласование средних арифметических по линейным значениям с получаемыми по ним значениями объема (площади). В данном случае надо пользоваться не средней арифметической, а *средней квадратической* для значений площади или *средней кубической* – для объемных мер. Вычисленная по линейным размерам средняя арифметическая даст меньшие значения объема и площади, чем те, которые будут получены по фактическому линейному параметру.

Вычисление средней квадратической и средней кубической невзвешенного ряда производят по формулам:

$$S = \sqrt{\frac{\sum X^2}{N}},$$

$$Q = \sqrt[3]{\frac{\sum X^3}{N}},$$

где средняя квадратическая – S, а средняя кубическая – Q

Если анализируются величины, представляющие собой изменения скорости процессов (например, прирост длины или диаметра побегов), либо величины, связанные с другими обратной пропорциональностью (число устьиц на единице площади листа, число клеток в поле зрения микроскопа, концентрация растворенного вещества в литре раствора и др.), следует использовать не среднюю арифметическую, а *среднюю гармоническую*, которая для невзвешенного ряда вычисляется как

$$H = \frac{N}{\sum (1/X)},$$

где H – средняя гармоническая, а остальные обозначения прежние.

Если необходимо обработать данные по темпам количественного изменения явлений (темпы роста растений, изменения численности популяций вредителей и др.), следует использовать *среднюю геометрическую*. Она вычисляется непосредственно по формуле либо путем логарифмирования:

$$G = \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \cdot X_4 \dots \cdot X_n}, \quad \lg G = \frac{\lg X_1 + \lg X_2 + \dots + \lg X_n}{n},$$

где G – средняя геометрическая.

Для приведенных выше средних параметров, вычисленных по переменным одной и той же совокупности, справедливо соотношение *правила мажоритарности*:

$$H \leq G \leq \bar{X} \leq S \leq Q.$$

Степень числовых различий между данными средними величинами тем больше, чем выше коэффициент вариации выборки.

Медиана – варианта, которая разделяет ряд на две равные по числу варианты части. Она используется как средний показатель описательного типа, а

близость ее величины к средней арифметической служит одним из доказательств близости данного распределения к симметричному. Вычисление медианты для взвешенного ряда проводят по формуле:

$$Me = X_e + C \left[\frac{0.5N - L}{f_e} \right],$$

где Me - обозначение медианы, X_e - начало класса, в котором находится медиана, C - величина классowego интервала, N - объем выборки, L - сумма частот классов, предшествующих классу, в котором находится медиана, начиная с меньшего. f_e - частота медианного класса.

Мода - это точка на оси абсцисс, которая соответствует максимальной частоте теоретической кривой распределения вариант. В обычных случаях за моду принимают наиболее часто встречающуюся варианту ряда. В ранжированном ряду за моду можно принять середину класса, который имеет наибольшую частоту. Положение моды относительно средней арифметической определяет степень скошенности (*асимметрии*) кривой распределения данного ряда, поскольку в нормальном распределении совпадают значения моды и средней арифметической. Мода вычисляется по формуле:

$$Mo = X_0 + C \left[\frac{f_2 - f_1}{2f_2 - f_1 - f_3} \right],$$

где Mo - мода, X_0 - начало модального класса, C - величина классowego интервала, f_1 - частота класса, предшествующего модальному, f_2 - частота модального класса, f_3 - частота класса, следующего за модальным. Наличие в одной выборке нескольких мод может указывать на неоднородность выборки.

Иногда для первичного анализа данных можно пользоваться зримым сопоставлением значений лимитов, средних показателей, медианы и моды у разных объектов, чтобы сравнить ряды по мере изменчивости. Так, анализируя данные биотестовой оценки летучих выделений некоторых травянистых видов в полевых условиях, мы с Л.Н.Волковой воспользовались для иллюстрации следующей формой графика (рис. 4).

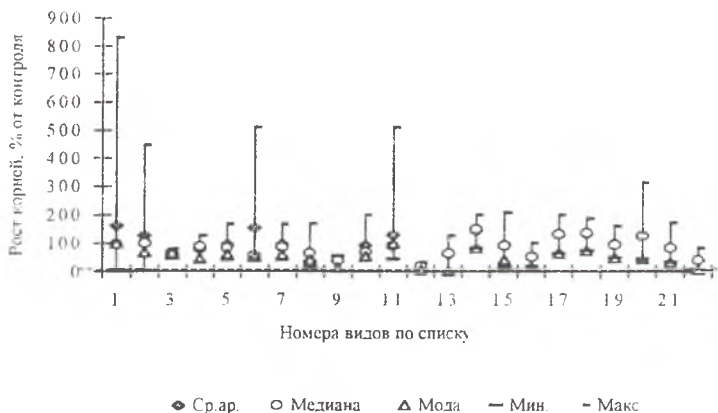


Рис. 4. Обобщенные результаты полевого биотирования активности летучих выделений ряда травянистых растений (из [11]).

Это позволило нам условно разделить объекты биотестирования на группы с различным характером влияния: виды 3, 9, 12 – активные ингибиторы биомассы, виды 4, 8, 13 – средние ингибиторы, виды 5, 21 – слабые ингибиторы, 7, 10, 15, 19 – практически не ингибирующие рост биомассы; обе группы видов можно было отнести к средне- и слабостимулирующим, наконец, виды 1, 2, 6, 11, 20 выделили в группу контрастных видов.

Завершая раздел, приведем два практических совета. Напомним читателю о том, что характер величины, для которой рассчитывается средняя арифметическая, определяет использование конкретных единиц измерения. Так, при измерении диаметра устьичного аппарата наиболее уместным будет выражать данные в микрометрах, а не в миллиметрах или более крупных единицах. Это избавит вас от усложненной записи чисел в виде десятичных дробей или отрицательных степеней 10.

Характер величины и сама процедура ее измерения определяют также число значащих цифр средней арифметической. Так, определяя длину корней проростков в миллиметрах с помощью обычной линейки, не стоит приводить итоговое значение средней арифметической в виде 12,326665 мм, как вы получили при обчете, – достаточно округлить до одного знака после запятой.

1.5.2. Меры рассеяния: лимиты, дисперсия, среднее квадратическое отклонение, коэффициент вариации, показатель точности опыта

Лимиты (размах варьирования) – это минимальное и максимальное значения признака в выборке. Они определяются при ранжировании ряда, и чем больше разность между минимальным и максимальным значениями ряда, тем признак изменчивее.

Дисперсия S^2 – мера рассеяния отдельных значений показателя вокруг среднего значения для этой совокупности. Этот показатель, как стандартное отклонение, является основной мерой рассеяния изучаемого признака. Она представляет собой частное от деления суммы квадратов отклонений на число измерений без единицы:

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{N - 1},$$

Среднее квадратическое (стандартное) отклонение σ представляет собой квадратный корень из дисперсии. Этот показатель, называемый также **сигма**, необходим и для характеристики выборки, и для вычисления некоторых других биометрических показателей:

$$\sigma = \bar{S} = \sqrt{S^2} = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N - 1}},$$

В выборках достаточно большого объема, которые находятся под влиянием множества разнообразных факторов, **диапазон изменчивости находится в пределах $\pm 3\sigma$ (правило трех сигм)**. При этом распределение соответствует нормальному, и выполняются следующие зависимости:

в области $\bar{X} \pm \sigma$ находится почти две третьих всех наблюдений, или 68,26%.

в области $\bar{X} \pm 2\sigma$ находится 95,46% всех значений переменной;

в области $\bar{X} \pm 3\sigma$ находится 99,73%, то есть практически все значения признака.

Сигма используется для расчета *ошибки средней арифметической (стандартной ошибки)*:

$$S_{\bar{x}} = \frac{\sigma}{\sqrt{N}} .$$

Цифровые значения результатов приводят в курсовых и дипломных работах, статьях, диссертациях в виде: $\bar{X} \pm S_{\bar{x}}$.

С помощью средней арифметической и ее ошибки вычисляют значение критерия Стьюдента для оценки достоверности средней арифметической:

$$t = \bar{X} / S_{\bar{x}} .$$

Сравнив полученное значение со стандартным (табличным) для числа степеней свободы $V = N - 1$, при $t > t_{\text{табл}}$, его считают достоверным для соответствующего уровня значимости (табл. 2 Приложения).

Показатель точности опыта Р выражает величину ошибки средней арифметической в процентах от средней:

$$P = \frac{S_{\bar{x}}}{\bar{X}} * 100\% ,$$

Точность опыта считают удовлетворительной, если величина показателя Р не превышает 5%, в противном случае необходимо увеличить число наблюдений или повторностей, а иногда необходимо воспользоваться иным методом исследования, чтобы повысить точность измерений. Существуют специальные методы, позволяющие установить число наблюдений, при котором будет достигнута определенная точность опытов – см., например, [5].

Коэффициент вариации V численно представляет собой отношение среднего квадратического отклонения к средней арифметической, выраженное в процентах:

$$V = \frac{\sigma}{\bar{X}} * 100\% ,$$

Коэффициент вариации может использоваться для сравнения изменчивости биологических показателей, которые принадлежат разным объектам наблюдений, относятся к различным процессам, сильно разнятся в количественном отношении и т.д.

Если коэффициент вариации не превышает 10%, показатель можно считать *жестким*, а изменчивость - незначительной. Изменения такого признака в ходе эксперимента весьма информативны. Если коэффициент вариации выше 10, но не превышает 20%, показатель *пластичен*, а изменчивость средняя. И наконец, если коэффициент вариации выше 20%, варьирование значительно, а признак *высоко пластичен*.

Г.Н.Зайцев [8], анализируя варьирование различных признаков у растений, предложил шкалу в процентах:

- 1) небольшое варьирование (от 0 до 4%);
- 2,3) нормальное варьирование (от 5 до 44%);
- 4) большое варьирование (от 45 до 64%);
- 5) очень большое варьирование (от 65 до 84%);

- 6) сверхбольшое (от 85 до 104 %);
7) аномальное (от 105 и больше).

1.5.3. Меры формы: коэффициенты асимметрии и эксцесса

Асимметрия (скошенность) – мера отклонения от нормального распределения, симметричного относительно максимальной ординаты. Степень асимметричности можно оценить простейшим способом, используя разность между средней арифметической и модой:

$$K_{AS} = \frac{(\bar{X} - Mo)}{\sigma},$$

где K_{AS} – коэффициент асимметрии, $\bar{X}$ – средняя арифметическая, Mo – мода, σ – среднее квадратическое отклонение. Для нормального распределения значение коэффициента равно 0. Положительное значение коэффициента соответствует положительной асимметрии (график распределения смещен влево, рис. 5А), отрицательное значение – отрицательной асимметрии (график смещен вправо, рис. 5В). Существует также вариант приведенной выше формулы, в котором медиану заменяет мода [2].

Второй способ для измерения степени асимметричности распределения использует отношение момента третьего порядка к кубу среднего квадратического отклонения:

$$AS = \frac{\sum (X_i - \bar{X})^3 f / N}{\sigma^3},$$

Полученная величина – **показатель асимметрии** AS – обычно сопровождается соответствующей ошибкой, которую определяют как

$$m_{AS} = \sqrt{6 / (N+3)}$$

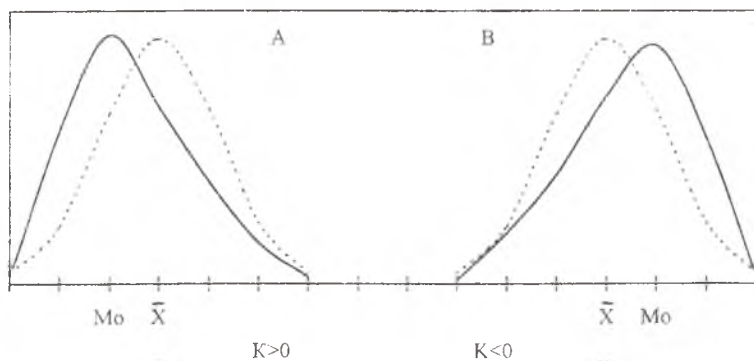


Рис. 5. Графики распределений с положительной (А) и отрицательной (В) асимметрией.

Примечание. Пунктиром показана теоретическая кривая нормального распределения.

Для того, чтобы оценить достоверность показателя асимметрии, вычисляют критерий Стьюдента, полученное значение которого сравнивают со стан-

дартным при числе степеней свободы, равном ∞ (табл.2 Приложения):

$$t = A_s / m_{A_s},$$

Исходя из табличных показателей, для значимой асимметрии величина t должна отвечать условию: $t \geq 1.96$.

Экссесс - показатель, позволяющий сравнить форму графика распределения с нормальным в области вершины, то есть оценить степень его остро- или туповершинности (вдавленности) (рис.6). При расчетах показателя эксцесса используют формулу:

$$E = \frac{\sum (X_i - \bar{X})^4 f}{N \cdot \sigma^4} - 3,$$

где E - показатель эксцесса, σ - среднее квадратичное отклонение, f - частоты вариантов, N - общее число вариантов.

Показатель отрицательного эксцесса («вдавленности») не может превысить -2, поскольку при этом его значения выборка состоит из вариант, относящихся к разным совокупностям, а провал на вершине кривой достигает оси абсцисс, разделяя график на две кривые распределения.

Положительный эксцесс по величине E не ограничен. Чем меньше значение E , тем ближе выборка к нормальному распределению, $|E| < 0.4$ соответствует незначительному эксцессу.

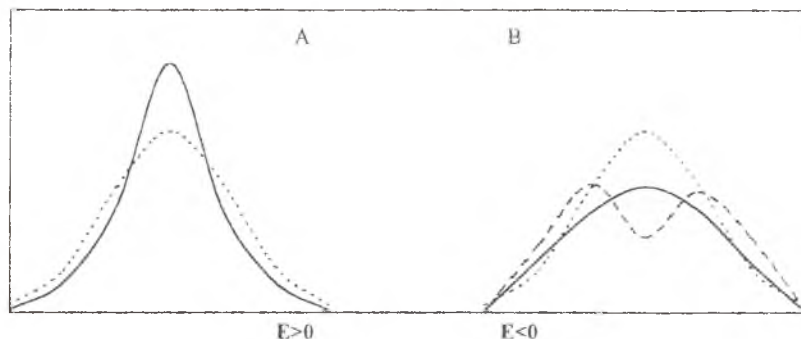


Рис.6. Графики распределений с положительным (А) и отрицательным (В) эксцессом.

Примечание. Мелким пунктиром показана теоретическая кривая нормального распределения.

1.6. Статистические показатели качественной изменчивости

Напомним, что к показателям качественной изменчивости будут принадлежать численно не измеряемые признаки растений — окраска лепестков, махровость цветка, наличие морфологических аномалий, присутствие повреждений насекомыми, заболеваний и пр. При анализе качественной изменчивости результаты исследований группируют, деля выборку на классы (группы) с различными качественными признаками.

Основные статистические показатели качественной изменчивости - *доля*

признака, показатель изменчивости, коэффициент вариации и ошибка выборочной доли.

Доля признака характеризует относительную частоту отдельной варианты в данной совокупности и обозначается $p_1, p_2, \dots, p_n$. Она может измеряться в долях единицы или процентах:

$$p_1 = n_1 / N,$$

и так далее, где $p_1, \dots, p_n$ - количество наблюдений, в которых присутствует - признак 1, ... признак n и т.д. Сумма всех долей в пределах совокупности ряда равна 1 (или 100%), в зависимости от способа выражения доли. При альтернативной изменчивости первый признак обычно обозначают p , второй - q , причем $p + q = 1$ (100%), соответственно доля второго признака $q = 1 - p$.

Показатель изменчивости качественного признака S позволяет оценить варьирование величин ряда относительно друг друга и определяется по формуле

$$S = \sqrt{pq},$$

В зависимости от соотношения p и q значение показателя изменяется от 0 до 0.5. Если количество градаций больше 2, формула принимает вид:

$$S = \sqrt{p_1 * p_2 * \dots * p_n}.$$

Коэффициент вариации качественных признаков - это показатель изменчивости, выраженный в процентах относительно максимальной изменчивости:

$$V_p = (S_p / S_{\max}) * 100\%,$$

где $S_{\max}$ - максимальная изменчивость.

Ошибка выборочной доли S - мера отклонения доли признака выборочной совокупности p от его доли во всей генеральной совокупности P . Данный показатель вычисляется по формуле:

$$S_p = S / \sqrt{N},$$

где S - показатель изменчивости качественного признака, N - объем выборки. Поскольку $S = \sqrt{pq}$, иное выражение для вычисления ошибки выборочной доли признака:

$$S_p = \sqrt{(p * q) / N}$$

Вероятность встречаемости признака p (или q) в интервале $p \pm S_p$ около 68%, в интервалах $p \pm 2S_p$ и $p \pm 3S_p$ - соответственно 95 и 99%.

ГЛАВА 2. СРАВНЕНИЕ ВЫБОРОК И СТАТИСТИЧЕСКИХ ПОКАЗАТЕЛЕЙ

2.1. Сравнение средних арифметических и дисперсий

Оценка достоверности различий между выборками - результатами опытов и наблюдений, - является весьма распространенной задачей для ботаников и экологов. Рассмотрим, как она решается.

Прежде, чем выполнять сравнение интересующих выборок, следует провести определенную подготовительную работу с данными. При количественном сравнении двух рядов (взвешенных или невзвешенных) необходимо вычислить для них средние арифметические, дисперсии или средние квадратические отклонения или ошибки средних арифметических. Необходимо знать объемы сравниваемых выборок, тип распределения вариант в них должен быть близок к

нормальному [8].

В самом простом случае можно сделать вывод о близости распределения к нормальному при совпадении между собой (близости) значений средней арифметической, медианы и моды.

Выборки можно сравнивать по средним арифметическим и степени вариабельности, причем обычно это делается одновременно.

2.1.1. Сравнение малых выборок

Если хотя бы одна из сравниваемых выборок мала ($N \leq 20$), то сначала оценивают различие дисперсий, затем одним из принятых способов сравнивают средние арифметические. Например, вычисляется критерий Фишера по формуле:

$$F = S_1^2 / S_2^2,$$

где числитель - большая дисперсия, а знаменатель - меньшая. Полученное значение критерия сравнивают с табличным F_t (табл.3 Приложений), для поиска которого используются числа степеней свободы $v_1 = N_1 - 1$, $v_2 = N_2 - 1$. N_1 и N_2 - число переменных у рядов с большей и меньшей дисперсией соответственно. Разница дисперсий достоверна, если полученное значение больше табличного при доверительном уровне $P = 95\%$. В таблице стандартных значений критерия Фишера приведены и значения критерия для доверительных уровней $P=99$ и $99,9\%$.

В случае несущественно различающихся или равных дисперсий средние арифметические сравнивают по формуле:

$$|t| = (\bar{X}_1 - \bar{X}_2) \sqrt{(1 - 2(N_1 - N_2)^{-1}) / (S_1^2/N_2 + S_2^2/N_1)}$$

t - обозначение критерия Стьюдента, знак его не принимается во внимание, а S_1^2 - соответствующая дисперсия. По таблице стандартных значений критерия Стьюдента (табл.3 Приложения) находят его значение при числе степеней свободы $v = N_1 + N_2 - 2$

Следует запомнить общий прием при использовании стандартных таблиц критериев: вычисленное по данным опыта (эмпирическое) значение следует сравнить по модулю со стандартным (табличным). Поиск нужного стандартного значения ведут, используя числа степеней свободы - они укажут вам строку (критерий Стьюдента) либо строку и столбец (критерий Фишера). При значении t , превосходящем табличное, разница между выборками достоверна для соответствующего доверительного уровня. Доверительный уровень, обозначаемый в виде $p = 95$ (99, 99,9)% либо как $p = 0,95$ (0,99, 0,999) означает, что проверяемая гипотеза верна на 95% из 100, а вероятность обратного события составляет 5%, и т.д. Так, если эмпирическое значение критерия Стьюдента превышает табличное при $p = 0,99$, сравниваемые выборки различаются с вероятностью 99%, - различия достоверны при уровне значимости 0,99, этот результат вы должны указать при обсуждении данных.

Если эмпирическое значение меньше табличного, различия между выборками недостоверны. Однако учтите: это не означает, что выборки имеют сходные показатели, не различаясь между собой, - вы всего лишь не подтвердили присутствия значимых различий математически. Возможная

причина этого – высокий уровень варьирования показателей внутри выборок. В этом случае может помочь увеличение числа измерений, то есть объема выборок, при проведении повторных опытов. Если это невыполнимо, то при обсуждении результатов вы можете говорить, например, о наличии тенденций к увеличению показателя А в первом варианте опыта по сравнению с контролем, и т.д.

В том случае, если дисперсии существенно различались, разница средних арифметических рассчитывается по формуле:

$$|t| = (\bar{X}_1 - \bar{X}_2) / \sqrt{S_1^2/N_2 + S_2^2/N_1}, \text{ или } |t| = (\bar{X}_1 - \bar{X}_2) / \sqrt{(S_{x1})^2 + (S_{x2})^2},$$

где S_1^2 и S_{x1} – дисперсия и ошибка соответствующего среднего арифметического. Число степеней свободы в этом случае вычисляется по формуле:

$$v = (N_1 + N_2 - 2) \cdot [0.5 + (S_1^2/S_2^2 + S_2^2/S_1^2)],$$

причем полученные результаты обычным способом округляют до целого значения

2.1.2. Сравнение больших выборок

Когда обе сравниваемые выборки велики, а их объемы слабо различаются или равны ($N > 20$), сравнение средних проводят по критерию Стьюдента по уже знакомой нам формуле:

$$|t| = (\bar{X}_1 - \bar{X}_2) / \sqrt{(S_{x1})^2 + (S_{x2})^2}$$

с числом степеней свободы $N = N_1 + N_2 - 2$.

Необходимо также указывать, для какого уровня значимости (доверности) делается вывод об обнаруженных различиях.

Если сравниваемые совокупности достаточно велики по объему и распределение близко к нормальному, можно сравнить средние, не прибегая к помощи таблиц и проведя более простые вычисления:

$$(\bar{X}_1 - \bar{X}_2)^2 / ((S_{x1})^2 + (S_{x2})^2) >_< 9,$$

то есть если левая часть неравенства больше 9, то различие средних арифметических значимо, а если полученный результат меньше 9, различие нельзя считать достоверным.

2.2. Сравнение коэффициентов вариации и показателей точности опыта

Ботаник и агроном нередко сталкиваются с необходимостью сравнения объектов по степени стабильности их показателей. Например, в условиях районов рискованного земледелия, к числу которых относится и Самарская область, устойчивый сорт со стабильными урожаями можно предпочесть более продуктивному, урожайность которого сильно колеблется по годам в зависимости от погодных условий. В подобных случаях сорта сравнивают по степени их варьирования с использованием коэффициентов вариации:

$$\frac{|v_1 - v_2|}{\sqrt{m^2_{v_1} - m^2_{v_2}}} > 3 + \frac{6}{N - 4},$$

где v_1 и m_{v_1} – соответственно коэффициент вариации и его ошибка. N – объем меньшей из сравниваемых выборок.

Если левая часть неравенства больше его правой части, различие коэффициентов вариации считается достоверным. Формула используется при числе вариантов N не меньше 5.

Примерно в тех же случаях, либо с целью контроля сравнения по коэффициентам вариации, используется сравнение показателей точности опыта. Формула аналогична приведенной выше:

$$\frac{|p_1 - p_2|}{\sqrt{m^2_{p_1} - m^2_{p_2}}} > 3 + \frac{6}{N - 4}.$$

Оценка результатов проводится так же.

2.3. Сравнение видового состава

При проведении исследований в различных регионах бывает необходимо сравнить видовой состав их флор либо фаун. Простые формулы, которые применяются для этой цели, позволяют вычислить коэффициенты K_1 – Жаккара и K_2 – Чекановского:

$$K_1 = \frac{C}{A + B - C}, \quad K_2 = \frac{2C}{A + B},$$

где A – число видов в первом регионе, B – во втором, C – число общих видов для двух сравниваемых регионов.

При сравнении флор (или фаун) большого числа районов составляют таблицы – *матрицы сравнения*, в которых приводят значения названных коэффициентов для пар сравниваемых объектов.

Используем в качестве иллюстрации результаты сравнения ряда фаунистических комплексов, проведенного С.А.Сачковым для Самарской Луки [12]. Здесь коэффициенты Жаккара выражены не в долях единицы, а в процентах:

Таблица 2

Матрица сходства лепидоптерокомплексов ландшафтов Самарской Луки (из [12])

Коэфф. Жаккара, %	Жигулевский	Винновский	Александровский	Переволоцко-Усинский	Рождественский	Шелехметский
Жигулевский	100	39	29	13	18	20
Винновский	39	100	54	28	36	35
Александровский	29	54	100	39	44	40
Переволоцко-Усинский	13	28	39	100	53	39
Рождественский	18	36	44	53	100	47
Шелехметский	20	35	40	39	47	100

Судя по данным таблицы, наибольшим фаунистическим сходством обла-дают Винновский и Александровский ландшафты, а также Рождественский и Переволокско-Усинский. С.А.Сачков объясняет полученные результаты тем, что первые два ландшафта весьма сходны биотопически и при различной залесенности содержат достаточно лесных и лесостепных стаций. Последние же ландшафты, ввиду высокой антропогенной освоенности, сближаются в результате преобладания синантропных и экономически значимых видов [12].

Использование коэффициентов Жаккара или Чекановского допустимо, если вы можете быть уверены в полноте полученных в ходе обследования флористических либо фаунистических списков.

Более математизированные приемы сравнения представлены в специ-альных разделах руководств по биометрии для ботаников и фаунистов [2, 13].

2.4. Использование критерия хи-квадрат в сравнении данных

Данный критерий – χ^2 – широко используется в биологических исследованиях при сравнении урожайности сортов, оценке эффективности химических препаратов, изучении расщепления потомства при скрещивании и др. В руководстве Г.Н.Зайцева [2] рассматриваются девять вариантов приме-нения этого критерия, причем указано, что все возможные способы его исполь-зования не исчерпаны. Перечислим некоторые ситуации, когда целесообразно вычисление критерия хи-квадрат:

- оценка совпадения теоретических и эмпирических частот взвешенных рядов (проверка соответствия фактических частот теоретической кривой распре-деления данного вида):

- сравнение дисперсий генеральной совокупности и выборки;
- сравнение двух эмпирических взвешенных рядов одинакового объема;
- сравнение двух эмпирических взвешенных рядов разного объема;
- изучение расщепления гибридов (совпадение ожидаемых и фактических численностей особей с определенными признаками).

Каждому из названных случаев соответствует свой вариант формулы для вычисления критерия. Приведем формулу для решения первой из названных задач:

$$\chi^2 = \sum \frac{(f - f')^2}{f'}$$

где χ^2 – критерий хи-квадрат, f – фактические частоты классов, f' – ожидаемые (теоретические) частоты.

Полученное значение критерия сравнивается с табличным (табл.4 При-ложения) для числа степеней свободы $v=k - 3$ при нормальном распре-делении, $v=k - 2$ при биномиальном и пуассоновском распределении, $v=k - 3$ при распределении Шарлье. Если вычисленное (эмпирическое) значение крите-рия меньше табличного, делается вывод, что теоретическое распределение несущественно отличается от эмпирического, то есть соответствует ему.

В качестве примера приведем данные Н.В.Прохоровой [14] отно-сительно распределения содержания тяжелых металлов в почвах Самарской

области (табл.3, рис.1 Приложения). Анализ большого цифрового материала позволил заключить, используя критерий хи-квадрат, что нормальному распределению подчиняется содержание в почвах марганца и титана – элементов в большей степени естественного происхождения

Таблица 3

Значения критерия хи-квадрат для теоретического и фактического распределения тяжелых металлов в почвах Самарской области (из [14])

Элемент	χ^2 фактич.	χ^2 теоретический			Число степ. свободы
		0.95	0.99	0.999	
Выборка по почвам Степного Заволжья, N=295					
Ti	21.17	23.68	29.14	36.12	14
Mn	21.40	19.68	24.72	31.26	11
Fe	21.71	15.51	20.09	26.12	8
Cr	182.10	19.68	24.72	31.26	11
V	758.10	21.03	26.22	32.91	12
Co	492.20	19.68	24.72	31.26	11
Ni	389.90	9.49	13.28	18.47	4
Cu	505.70	7.82	11.34	16.27	3
Zn	300.60	11.07	15.09	20.52	5
Si	4.81	14.07	18.48	24.32	7
Rb	89.84	24.99	30.57	37.69	13
Pb	615.50	16.92	21.67	27.88	9

Интересно также отметить, что при анализе весьма разнообразных данных (по почвам Самарской области в целом) автор практически не обнаруживает нормального распределения ни для одного элемента. При анализе данных в пределах одного региона – Степного Заволжья – диапазон варьирования показателей, очевидно, сужается, и ряд элементов демонстрирует соответствие нормальному распределению.

ГЛАВА 3. КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

3.1. Корреляционная связь биологических показателей

Между показателями биологических объектов может существовать *функциональная связь*, при которой одному значению признака X, принятого за аргумент, соответствует только одно значение признака функции Y. Но чаще одному значению аргумента может соответствовать не одно, а несколько значений функции. При этом обнаруживается корреляционная связь той или иной степени тесноты (силы). Ввиду неоднородности объектов ботанических исследований и воздействия на них целого ряда разнообразных факторов связи, обнаруживаемые ботаниками и экологами в экспериментах, представлены почти исключительно корреляционными. Между признаками может иметь место *прямая корреляционная связь* (большая высота растения - большая длина его корня и т.д.) и *обратная корреляционная связь* (чем больше масса растения, тем меньше число растений, которые могут существовать на единице площади).

Известны случаи, когда одни и те же признаки, в зависимости от уровня

их числовых значений, может связывать прямая либо обратная корреляция. Так, число соцветий и высота растений гелениума находятся в прямой связи, но при увеличении роста растений выше некоего уровня связь с числом соцветий на растениях становится обратной [2].

Сложность изучения корреляции вытекает из взаимосвязи, в той или иной степени, всех признаков растительного организма. Метод парной корреляции, который мы опишем, позволит рассмотреть связь именно между двумя признаками, которая имеет достаточно ясный биологический смысл. В противном случае рассчитанная математически корреляция покажет связь между двумя признаками, не зависящими друг от друга, но подчиняющимися изменениям третьего признака или совокупности других признаков.

В качестве упрощенного способа первоначальной оценки корреляции пары рядов можно предложить следующий графический способ. Для малочисленных невзвешенных рядов признаков принимают данные ряда 1 за значения X, ряда 2 - за значения Y для N точек с координатами (X,Y). В осях координат с соответствующим масштабом наносят точки и по характеру их рассеяния делают предварительные выводы о возможной связи между исследуемыми параметрами (рис. 7).

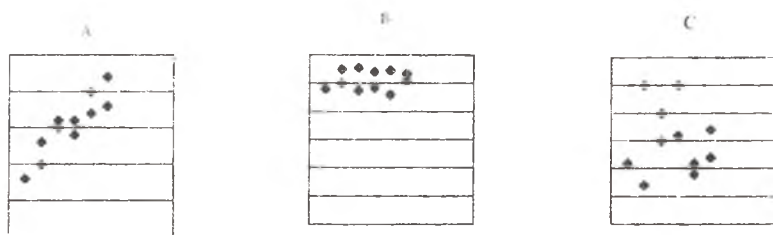


Рис. 7. Различные формы корреляционных связей между показателями X и Y

Так, тип А, вероятно, свидетельствует о наличии достаточно тесной прямой связи между изучаемыми показателями; тип В - о весьма вероятном отсутствии связи между данными признаками; тип С - о наличии связи, которая не относится к парным корреляциям и не исчерпывается взаимодействием двух взятых показателей.

3.2. Коэффициент корреляции в малых выборках

Рассматриваемый вид корреляции может в различных руководствах носить названия простой линейной корреляции. Основной ее показатель – коэффициент корреляции r , который вычисляют по формуле:

$$r = \frac{\sum XY - (\sum X * \sum Y) / N}{\sqrt{[\sum X^2 - (\sum X)^2 / N] * [\sum Y^2 - (\sum Y)^2 / N]}}$$

где X и Y- варианты двух анализируемых рядов. N - число пар вариантов (X, Y) в рядах.

Для оценки существенности коэффициента корреляции r вычисляют его ошибку S_r и критерий существенности t_r :

$$S_r = \sqrt{(1 - r^2)/(N - 2)}, \quad t_r = r/S_r.$$

Вычисленное значение t_r сравнивают с табличным значением критерия Стьюдента для уровней $p=95$ и 99% и числе степеней свободы $v=N-2$. Если опытное значение превышает табличное, корреляционная связь существенна.

Если коэффициент корреляции оказался недостоверным, это может быть объяснено по крайней мере двумя причинами: малой выборкой или отсутствием исследуемой связи в генеральной совокупности данного объекта. Некоторое снижение величины коэффициента r может произойти из-за излишнего округления значений X и Y . Минимальные значения коэффициента корреляции r , существенные при данном числе вариант, приведены в соответствующем разделе руководства Г.Н.Зайцева [2]. Согласно расчетам, выполненным при построении этой таблицы, для суждения о достоверности $r = 0,32$ требуется, чтобы в ряду было не менее 38 пар вариант, и так далее.

Коэффициент корреляции r - безразмерная величина, $|r| < 1$.

Знак "-" соответствует отрицательной корреляционной связи, когда рост значений показателя А коррелирует со снижением показателя В. При $r = 0$ связь отсутствует, при $r = 1$ - превращается в функциональную (линейную), что характерно для взаимообусловленных величин. По величине коэффициента корреляции судят о тесноте связи. $r < 0,3$ - связь слабая, $0,3 < r < 0,7$ - связь средняя, $r > 0,7$ - связь сильная.

При анализе парных связей между несколькими показателями полученные результаты оформляют в виде корреляционной матрицы, внешне напоминающей матрицу сходства (см. раздел 2.3.). Приведем, например, результаты корреляционного анализа по данным содержания фотосинтетических пигментов в листьях березы:

Таблица 4

Результаты корреляционного анализа некоторых показателей обмена веществ в листьях березы бородавчатой, г. Самара, лето 1997

	Район	Месяц	Хлорофилл А	Хлорофилл В	Каротиноиды	Фенолы
Район	1,00	0,00	-0,03	-0,01	-0,10	-0,18
Месяц	0,00	1,00	-0,49	-0,43	-0,42	0,66
Хлорофилл А	-0,03	-0,49	1,00	0,98	0,91	-0,79
Хлорофилл В	-0,01	-0,43	0,98	1,00	0,88	-0,76
Каротиноиды	-0,10	-0,42	0,91	0,88	1,00	-0,65
Фенолы	-0,18	0,66	-0,79	-0,76	-0,65	1,00

При использовании матричной формы диагональ (значения $r = 1$) оплит

таблицу на две одинаковые половины, опустимо не заполнять одну из них, как иногда бывает.

3.3. Коэффициент корреляции в больших выборках

При больших группах данных ($N > 100$) нахождение самого коэффициента r может производиться по приведенной в разделе 3.2. формуле, либо пользуются при-годной для малых выборок формулой:

$$r = \frac{\sum(X - \bar{X}) * \sum(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 * \sum(Y - \bar{Y})^2}}.$$

Изменится характер вычисления ошибки коэффициента корреляции: для рядов с $N > 100$ она вычисляется как $S_r = (1 - r^2) / \sqrt{n}$, а t будет рассчитываться как указано выше. Оценка достоверности r будет проводиться аналогично случаю с малыми выборками.

3.4. Коэффициент и уравнение регрессии

После того, как выяснен вопрос о наличии корреляционной связи между исследуемыми признаками, можно приступить к оценке закономерностей количественного изменения одного из признаков (функции) при изменении другого (аргумент). Для этого пользуются коэффициентом регрессии и уравнением регрессии, при необходимости используют и графическое выражение связи, вычисляя по найденному уравнению точки теоретической кривой зависимости Y от X .

Коэффициент регрессии b_{yx} показывает, в каком направлении и на какую величину в среднем изменяется функция Y при изменении аргумента X на единицу измерения. Коэффициент регрессии имеет тот же знак, что и соответствующий коэффициент корреляции, и вычисляется как

$$b_{yx} = \sum(X - \bar{X}) * \sum(Y - \bar{Y}) / \sum(X - \bar{X})^2$$

Возможен и расчет обратной связи, когда аргумент и функция меняются местами и формула приобретает вид

$$b_{xy} = \sum(X - \bar{X}) * \sum(Y - \bar{Y}) / \sum(Y - \bar{Y})^2$$

Для данных коэффициентов также вычисляют ошибку S_b и критерий существенности t_b :

$$S_{byx} = S_r \sqrt{\sum(X - \bar{X})^2 / \sum(Y - \bar{Y})^2}, \quad t_b = b / S_b$$

При этом выполняются зависимости: $t_b = t_r$, $b_{yx} * b_{xy} = r^2$, что можно использовать при проверке расчетов.

Уравнение регрессии представляет собой математическую формулу для данной корреляционной связи. Рассмотрим простейший случай, когда между изучаемыми признаками установлено наличие простой линейной корреляции, уравнение которой имеет вид: $Y = A * X + B$.

В биометрии уравнение регрессии, соответствующее простой линейной корреляции, иногда представляют в виде: $Y = Y + b_{yx}(X - \bar{X})$.

В этом случае для построения теоретической линии регрессии в формулу подставляют значения X , Y , b , после этого берут два экстремальных значения $X - X_{\min}$ и $X_{\max}$, вычисляют соответствующие им значения Y и строят

по ним линию регрессии в соответствующих координатных осях.

3.5. Линейное уравнение калибровочного графика

Уравнением в форме $Y = A \cdot X + B$ мы советуем пользоваться, когда в работе применяется какой-либо колориметрический метод определения концентрации вещества, активности фермента и т.д.

Чаще всего в этом случае рекомендуют использовать *калибровочный график*, который построен по данным оптической плотности (Y) и концентрации вещества (X) стандартных растворов на миллиметровой бумаге и позволяет непосредственно находить концентрацию вещества в растворе по известной оптической плотности. Однако нередко возникают проблемы с масштабом калибровочного графика, определение носит оттенок субъективности и достаточно утомительно для глаз.

В то же время указанная зависимость представляет собой линейное уравнение, которое по тем же данным стандартных растворов можно с легкостью решить. В конкретном случае оно примет вид: $Y = A \cdot X$, коэффициент A будет равен: $A = Y/X$ для конкретной точки, а для графиков в целом мы предлагаем вычислять его как $A = (Y_1/X_1 + Y_2/X_2 + \dots + Y_N/X_N) / N$, то есть используя вычисление среднего арифметического.

Для проверки рекомендуем построить график по найденному уравнению и в тех же координатных осях нанести фактические точки (данные стандартных растворов). Если разброс точек идет вокруг теоретической линии, можно пользоваться проверяемым уравнением для нахождения концентрации исследуемого вещества (Y) по найденной в опыте оптической плотности (X).

ГЛАВА 4. ДИСПЕРСИОННЫЙ АНАЛИЗ

4.1. Дисперсионный анализ и его биологический смысл

Изменчивость признаков живых организмов определяется влиянием многочисленных факторов. Степень их влияния неодинакова, а сами факторы могут быть *организованными (регулируемыми) и случайными*.

Первые представляют действие изучаемого фактора, которому в эксперименте придается несколько значений (*уровней, градаций*), в соответствии с градациями признака выделяется несколько вариантов опыта.

Вторые обуславливаются естественным варьированием любых признаков биологических объектов, это не контролируемые в данном опыте факторы, оказывающие случайное влияние.

Относительная роль отдельных факторов в общей изменчивости исследуемого признака характеризуется *дисперсией* и может быть изучена путем *дисперсионного анализа или анализа рассеяния*. Его смысл заключается в изучении статистического влияния одного или нескольких факторов на результативный признак, установлении роли отдельных факторов в изменчивости изучаемого признака и расчленении общей дисперсии на составляющие, вызванные влиянием отдельных факторов или их групп [5,6].

Разработка методов дисперсионного анализа была проведена английским статистиком Р.А. Фишером на Ротэмстедской агробиологической станции. мето-

ды предполагали изучение вклада агробиологических факторов (сорт, удобрения, условия роста и т.п.) в урожайность и иные свойства растений [4]. В англоязычной литературе метод имеет название ANOVA (Analysis of Variance).

При проведении дисперсионного анализа оценивают варьирование по вариантам, повторностям, остаточное варьирование внутри групп и общее варьирование результирующего признака в опыте. Иными словами, *дисперсионный анализ - это расчленение общей суммы квадратов отклонений и общего числа степеней свободы на части, соответствующие структуре эксперимента, и оценка значимости действия и взаимодействия изучаемых факторов по F-критерию.*

При этом одновременно обрабатывают несколько выборок (вариантов) единого статистического комплекса, представляющего собой совокупность градаций с измеренными значениями исследуемого показателя.

По мнению Н.А.Плохинского [6], основным показателем силы влияния следует считать отношение факториальной суммы квадратов к общей сумме квадратов. На основе этого показателя вычисляют ошибку показателя силы влияния, которую используют для определения доверительных границ основного показателя.

4.2. Схема однофакторного дисперсионного анализа

Для изложения схемы проведения анализа мы использовали руководства Г.Н.Зайцева [2] и Н.А.Плохинского [6], к которым и отсылаем читателя при необходимости углубленного освоения данного метода.

Удобно вести записи и вычисления в форме таблиц.

Первая таблица представляет алгоритм вспомогательных вычислений, которые позволят перейти к расчету необходимых статистик. В ходе вычислений определяют суммы данных по повторностям и вариантам, суммы квадратов и квадраты сумм (табл.5). Рассмотрим форму таблицы, как ее представляет Г.Н.Зайцев [2].

Таблица 5
Суммы квадратов однофакторного дисперсионного анализа (рабочая таблица 1)

Градация фактора (n)	Повторности (g)			S_i	$\sum_i X^2$	S_i^2	M_i
	I	II	III				
Контроль							
Вариант 1							
Вариант 2							
Вариант 3							
S_2							
$\sum_i X^2$				$\sum X$		$\sum S_i^2$	
S_2^2							
M_i				$\sum X^2$		$\sum S_2^2$	

Примечание. Стрелки в таблице указывают направление суммирования при вычислении обозначенных величин – сумм, сумм квадратов, квадратов сумм.

На первом этапе задачей однофакторного дисперсионного анализа является расчленение общего (суммарного) варьирования на составные части:

$$\Theta = \Theta_1 + \Theta_2 + \Theta_3,$$

где Θ - сумма квадратов отклонений по общему варьированию данных, Θ_1 - сумма квадратов отклонений по градациям фактора, Θ_2 - сумма квадратов отклонений по повторностям опыта, Θ_3 - сумма квадратов отклонений по остаточному варьированию данных опыта.

Соотношение чисел в этой сумме будет соответствовать вкладу того или иного эффекта в общее разнообразие. Так, абсолютное преобладание величины слагаемого 1 укажет на существенное влияние фактора, слагаемого 2 - на высокую степень неоднородности данных в вариантах, слагаемого 3 - на то, что варьирование могло быть вызвано действием не учитывавшихся экспериментатором, но серьезно повлиявших факторов.

Общую сумму квадратов отклонений опытных данных вычисляют по формуле:

$$\Theta = \Sigma X^2 - (\Sigma X)^2 / N,$$

где N - общее число наблюдений, оно равно $N = n \cdot g$ (произведению числа градаций опыта на число повторностей).

Сумму квадратов отклонений по градациям опыта (факториальную) вычисляют как

$$\Theta_1 = [\Sigma S_1^2 - (\Sigma X)^2 / n] / g.$$

Сумма квадратов отклонений по повторностям опыта вычисляется по аналогичной формуле:

$$\Theta_2 = [\Sigma S_2^2 - (\Sigma X)^2 / g] / n.$$

Сумма квадратов отклонений по остаточной вариации вычисляется согласно равенству:

$$\Theta_3 = \Theta - \Theta_1 - \Theta_2.$$

Далее нужно вычислить числа степеней свободы для соответствующих сумм квадратов

По общей сумме квадратов отклонений число степеней свободы равно числу всех наблюдений без единицы: $V = N - 1$.

По факториальной сумме квадратов число степеней свободы равно числу градаций без единицы: $V_1 = n - 1$, по повторностной сумме квадратов оно равно числу повторностей без единицы: $V_2 = g - 1$. Число степеней свободы для остаточной суммы квадратов отклонений равно $V_3 = V - V_1 - V_2$.

Далее вновь обращаемся к форме таблицы.

Таблица 6

Дисперсионный анализ данных (рабочая таблица 2)

Варьирование данных	Сумма квадратов	Степень свободы	Дисперсия. $\sigma^2 = \theta / v$	Критерий Фишера	
				F	F табл. $p=0,95$
Общее	Θ	V	-	-	-
Факториальное	Θ_1	V_1	-	F_1	-
Повторностное	Θ_2	V_2	-	F_2	-
Остаточное	Θ_3	V_3	-	-	-

Для вычисления критерия F (Фишера) полученные факториальную и повторностную дисперсии сравнивают с остаточной, при этом всегда делят боль-

шую дисперсию на меньшую:

$$F_1 = \sigma_1^2 / \sigma_{\text{ост}}^2, \quad F_2 = \sigma_{\text{ост}}^2 / \sigma_2^2$$

Вычисленный критерий F (Фишера) сравнивают с табличным значением (табл.3 Приложения) при определенном уровне значимости p . Поиск нужного столбца и строки проводят, используя числа степеней свободы. Если вычисленное факториальное значение критерия Фишера превосходит табличное, это соответствует достоверному (при данном уровне значимости) влиянию фактора на исследуемый показатель. Если же факториальное значение критерия Фишера ниже табличного, это может быть следствием малого числа повторностей либо высокой вариабельности данных. Превышающее табличное повторностное значение критерия Фишера указывает на достоверное варьирование внутри вариантов.

Ошибка общей средней арифметической определяется по формуле

$$m_m = \sqrt{\sigma_{\text{ост}}^2 / N}$$

Аналогичным образом вычисляют ошибку для частных средних арифметических. Для оценки достоверности различий между частными средними арифметическими наиболее простым способом является применение критерия Стьюдента (см. сравнение средних арифметических). Для вариантов опыта вычисляют также показатели точности опыта По Н.А.Плохинскому [6]. отношение факториальной суммы квадратов k общей дает *показатель силы влияния фактора*:

$$\eta_2 = \theta_1 / \theta$$

Ошибка данного показателя определяется по формуле

$$m_{\eta} = (1 - \eta^2) * [(g-1) / (N-g)]$$

Предложено использовать эту ошибку для определения доверительных границ генерального параметра:

$$\bar{\eta}^2 = [(\eta^2 - \Delta) + (\eta^2 + \Delta)], \text{ где } \Delta = F_{\text{табл.}} * m_{\eta}$$

Например, если показатель силы влияния, вычисленный в опыте, составил 0,568, а его ошибка – 0,069, то при значении критерия Фишера ($p = 0.95$) доверительные границы показателя силы влияния составят 0,38 и 0,76. Это позволяет сделать вывод, что влияние изученного фактора может составить ($p=0,95$) не менее 38% и не более 76% от общего влияния всей суммы факторов.

Более сложным, но в принципе аналогичным методом является проведение дву- и многофакторного дисперсионного анализа. Однако, как вполне понятно, сама процедура расчетов становится более громоздкой и оставляет желать применения вычислительной техники.

ГЛАВА 5. НЕКОТОРЫЕ ЗАМЕЧАНИЯ ОБ ИСПОЛЬЗОВАНИИ ПЭВМ

5.1. Использование программ на языке BASIC

Программирование на языке BASIC в принципе изучается в школьном курсе информатики. Однако данный язык программирования достаточно прост и логичен, чтобы им мог воспользоваться даже студент-биолог, изучивший достаточные руководства, например, [15]. На этом языке удобно составлять сравнительно небольшие программы, позволяющие решить конкретные, сиоминутные задачи биолога:

- обсчет первичных данных, которые не требуют хранения в виде матриц.
- вычисление показателей, находящихся в зависимости от ряда иных величин, например, вычисление содержания вещества в образце по данным калибровочного графика и оптической плотности раствора и др.

В качестве примера предлагаем ознакомиться с составленными программами на языке BASIC, вынесенными в приложения. Первая из них позволяет вычислить основные статистические показатели для невзвешенного ряда, и составлена нами в соответствии с формульной основой руководства Г.Н.Зайцева [2].

Вторая программа делает возможным применение сравнительно мало используемого метода сравнения двух процессов – метода Н.А.Плохинского. Этот метод по сути является вариантом двухфакторного анализа и применим в случаях, когда необходимо сравнить протекание двух процессов и определить достоверность различия между ними в качественном и количественном отношении – по критериям непараллельности и среднего уровня соответственно[6].

Первый фактор имеет в данном случае две градации – принадлежность объектов к одному из двух сравниваемых сортов, видов, категорий, второй же может иметь несколько градаций (например, по числу доз тестируемого соединения, включая и контрольный вариант).

Критерий непараллельности позволяет оценить достоверность динамики реакций двух объектов на данный фактор, критерий среднего уровня – достоверность количественных различий показателей, принадлежащих двум сравниваемым объектам.

По составлению программ на языке BASIC хочется сделать лишь одно замечание: главное, что необходимо вам, это хорошо разобраться в математической основе будущей программы, проанализировать формулы и грамотно перевести их в форму операторов программы. Внимание и тщательность обеспечат успех работы. Однако без глубокой подготовки в области информатики вряд ли стоит браться за выполнение слишком обширных задач в этой области.

5.2. Использование пакетов прикладных программ

Современная вычислительная техника значительно облегчает работу биолога-исследователя, освобождая его от рутинных расчетов. Тем не менее неперменным условием является понимание задач, которые ставятся при проведении математической обработки данных, и биологического смысла, заключенного в получаемых показателях. Использование пакетов прикладных программ практически сводит работу биолога к следующим операциям:

- подготовка первичной информации для математической обработки, введение данных в виде матриц (формирование баз данных),
- выбор процедуры математической обработки, задание ее условий,
- осмысление и биологическая интерпретация результатов, представление их в одной из графических форм (последнее вновь осуществляется с помощью компьютера).

Таким образом, главное, что следует в минимуме уметь, - грамотно поставить задачу и оценить полученную информацию, вся же техническая работа будет выполнена компьютером.

В этом разделе мы лишь упомянем о двух пакетах программ, которые можно использовать в работе биолога.

Первая из них – *STATGRAPHICS*. Этот пакет используется давно и в различных версиях. Он позволяет решить широкий круг задач [16,17], однако не лишен недостатков. К ним можно отнести ограниченные возможности используемой графики, а также англоязычность, что затруднит работу с ним при слабом знании языка. Однако *STATGRAPHICS* вполне применим для работы с экспериментальными данными на ЭВМ без использования Windows.

В последнее время более широкое распространение получает использование *Excel*, причем данный пакет программ представляет широчайшие возможности как в обработке данных, так и в их графической интерпретации. Однако в этом случае требуется достаточно хороший навык работы в системе Windows. Поскольку руководств по Windows и *Excel* выпускается сейчас очень много, приведем хотя бы один из источников – на наш взгляд, достаточно простой в освоении [18].

И последний совет читателю: пусть математика при всей ее привлекательности никогда не заслоняет для вас биологической сущности процессов и показателей, когда вы поверяете гармонию жизни алгеброй!

Литература

1. Леонов В.П., Ижевский П.В. Об использовании прикладной статистики при подготовке диссертационных работ по медицинским и биологическим специальностям // Бюллетень ВАК. 1997. № 4. С.12-17.
2. Зайцев Г.Н. Математическая статистика в экспериментальной ботанике. М.: Наука, 1984. 424 с.
3. Лакин Г.Ф. Биометрия. М.: Высшая школа, 1980. 352 с.
4. Леонов В.П. Обработка экспериментальных данных на программируемых микрокалькуляторах. Томск: Изд-во Том. ун-та, 1990. 376 с.
5. Пересыпкин В.Ф., Коваленко С.Н., Шелестова В.С., Асатур М.К. Практикум по методике опытного дела в защите растений. М.: Агропромиздат, 1989. 175 с.
6. Плохинский Н.А. Алгоритмы биометрии. М.: Изд-во Моск. ун-та, 1980. 150 с.
7. Мамаев С.А., Махнев А.К. Проблемы биологического разнообразия и его поддержания в лесных экосистемах // Лесоведение. 1996. № 5. С.3 - 10.
8. Быков Б. А. Экологический словарь, 2-е изд. Алма-Ата: Наука. 1988. 212 с.
9. Горчаковский П.Л., Зуева В.Н. Возрастная структура и динамика малых изолированных популяций уральских эндемичных астрагалов // Экология. 1984. С. 3-11.
10. Алексеев В.А. Диагностика жизненного состояния деревьев и древостоев // Лесоведение. 1989. N 4. С.51-57.
11. Волкова Л.Н., Кавеленова Л.М. Активность летучих выделений дикорастущих травянистых растений в полевых условиях // Вопросы экологии и охраны природы в лесостепной и степной зонах. Самара: Изд-во Самарского госуниверситета, 1996. С.149-158.
12. Сачков С.А. Эколого – фаунистический обзор чешуекрылых (*Lepidoptera*) Самарской Луки: Дисс. ... канд. биол. наук. СПб, 1996. 290 с. (Рукопись).
13. Песенко Ю.А. Принципы и методы количественного анализа в фаунистических исследованиях. М.:Наука, 1982. 288 с.
14. Прохорова Н.В. Распределение тяжелых металлов в почвах и растениях в зависимости от экологических особенностей лесостепного и степного Поволжья (на примере Самарской области): Дисс... канд. биол. наук. Самара. 1996. 289 с. (Рукопись).
15. Дьяков В.П. Справочник по алгоритмам и программам на языке бейсик для персональных ЭВМ. М.: Наука, 1987. 240 с.
16. Григорьев С.Г., Левацковский В.В., Перфилов А.М., Юнкеров В.И. Пакет прикладных программ STATGRAPHICS на персональном компьютере: Практическое пособие по обработке результатов медико-биологических исследований. СПб, 1992. 104 с.
17. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере / Под ред. В.Э.Фигурнова. М.: Инфра-М, Финансы и статистика, 1995. 384 с.
18. Тофель Т. Excel 7.0 для Windows 95. Секреты и советы / Пер. с нем. М.: Восточная Книжная Компания, 1996. 336 с.

ПРИЛОЖЕНИЯ

Таблица 1П

Значения критериев V_1 и V_N на уровнях значимости
0,99 (первое число) и 0,95 (второе число)

N	$V_{N(1)}$	N	$V_{N(1)}$	N	$V_{N(1)}$
4	0,991 0,955	13	0,520 0,410	22	0,414 0,320
5	0,916 0,807	14	0,502 0,395	23	0,407 0,314
6	0,805 0,689	15	0,486 0,381	24	0,400 0,309
7	0,740 0,610	16	0,472 0,369	25	0,394 0,304
8	0,683 0,554	17	0,460 0,359	26	0,389 0,299
9	0,635 0,512	18	0,449 0,349	27	0,383 0,295
10	0,597 0,477	19	0,439 0,341	28	0,378 0,291
11	0,566 0,450	20	0,430 0,334	29	0,374 0,287
12	0,541 0,428	21	0,421 0,327	30	0,369 0,283

Таблица 2П

Значения критерия Стьюдента при различных уровнях значимости

Число степ. свободы	$\alpha, \%$			Число степ. своб.	$\alpha, \%$		
	5	1	0,1		5	1	0,1
1	12,71	63,66	64,60	18	2,10	2,88	3,92
2	4,30	9,92	31,60	19	2,09	2,86	3,88
3	3,18	5,84	12,92	20	2,09	2,85	3,85
4	2,78	4,60	8,61	21	2,08	2,83	3,82
5	2,57	4,03	6,87	22	2,07	2,82	3,79
6	2,45	3,71	5,96	23	2,07	2,81	3,77
7	2,37	3,50	5,41	24	2,06	2,80	3,75
8	2,31	3,36	5,04	25	2,06	2,79	3,73
9	2,26	3,25	4,78	26	2,06	2,78	3,71
10	2,23	3,17	4,59	27	2,05	2,77	3,69
11	2,20	3,11	4,44	28	2,05	2,76	3,67
12	2,18	3,05	4,32	29	2,05	2,76	3,66
13	2,16	3,01	4,22	30	2,04	2,75	3,65
14	2,14	2,98	4,14	40	2,02	2,70	3,55
15	2,13	2,95	4,07	60	2,00	2,66	3,46
16	2,12	2,92	4,02	120	1,98	2,62	3,37
17	2,11	2,90	3,97	00	1,96	2,58	3,29
P	0,05	0,01	0,001	P	0,05	0,01	0,001

Таблица 3П

Значения критерия Фишера при уровнях значимости
 $p=0.95$ (верхняя строка) и $p=0.99$ (нижняя строка)

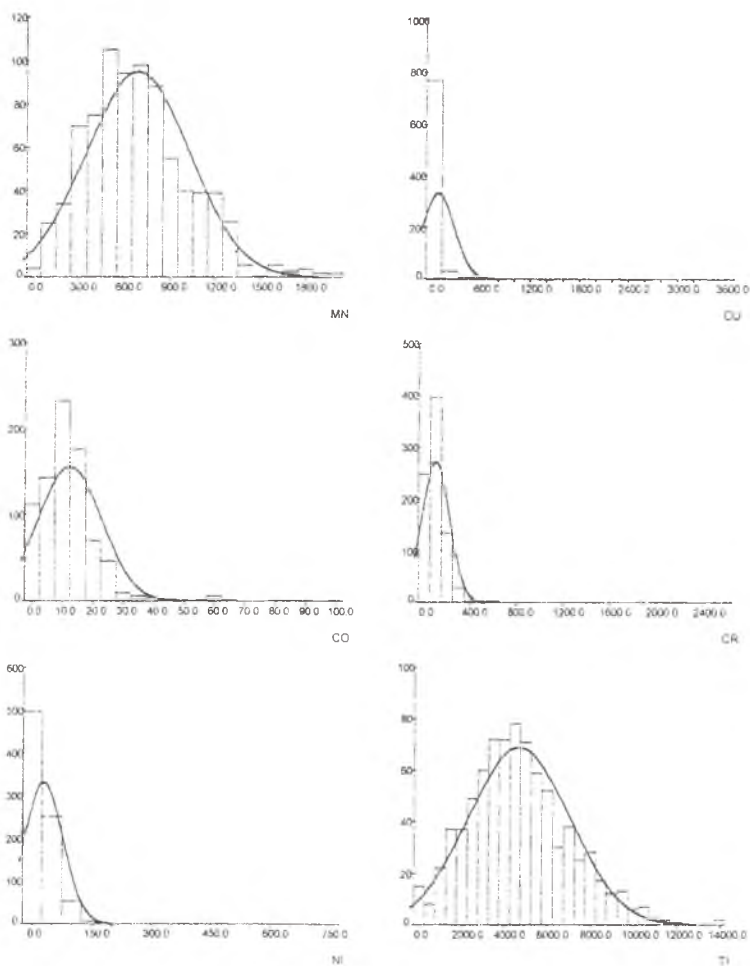
V_2	V_1 (степени свободы для большей дисперсии)							
	1	2	3	4	5	6	7	8
1	161	200	216	225	230	234	237	239
	4052	4999	5403	5625	5764	5859	5928	5982
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37
	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85
	34.12	30.82	29.16	28.71	28.42	27.91	27.67	27.49
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	6.00
	21.20	18.00	16.89	15.98	15.52	15.21	14.98	14.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82
	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15
	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73
	12.25	9.55	8.47	7.85	7.46	7.19	6.99	6.84
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44
	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03

Таблица 4 П

Значения критерия хи-квадрат

Число степ. сво- боды V	Уровни значимости			Число степ. сво- боды V	Уровни значимости		
	0.05	0.01	0.001		0.05	0.01	0.001
1	3.841	6.635	10.827	16	26.296	32.000	39.252
2	5.991	9.210	13.815	17	27.587	33.409	40.790
3	7.815	11.345	16.268	18	28.869	34.805	42.512
4	9.488	13.277	18.465	19	30.144	36.191	43.820
5	11.070	15.086	20.517	20	31.410	37.566	45.315
6	12.592	16.812	22.457	21	32.671	38.932	46.797
7	14.067	18.475	24.322	22	33.924	40.289	48.268
8	15.507	20.090	26.125	23	35.172	41.638	49.728
9	16.919	21.666	27.877	24	36.415	42.980	51.179
10	18.307	23.209	29.588	25	37.652	44.314	52.620
11	19.675	24.725	31.264	26	38.885	45.642	54.052
12	21.026	26.217	32.909	27	40.113	46.963	55.476
13	22.362	27.688	34.528	28	41.337	48.278	56.893
14	23.685	29.141	36.123	29	42.557	49.588	58.302
15	24.996	30.578	37.697	30	43.773	50.892	59.703

Приложение 5
Рис. 1П (из [14])



Теоретическое (кривые) и фактическое (столбцы) распределение показателей содержания тяжелых металлов в почвах Степного Заволжья Самарской области (общий массив данных, $n=295$): по вертикали - частота, по горизонтали - содержание элемента, мг/кг.

Программа для обработки невзвешенного ряда

```

1 DIM X(250). Y(250)
3 PRINT "Название ряда?"
5 INPUT A$
7 PRINT "Число данных в выборке?"
10 INPUT N
13 PRINT "Введите данные, по одному после знака вопроса"
15 FOR I = 1 TO N
17 INPUT A
20 X(I) = A
23 NEXT I
25 PRINT "Контроль или опыт обрабатывается?"
27 PRINT "Введите 0 для контроля, 1 для опытного ряда"
30 INPUT K1
32 IF K1 = 0 GOTO 45
35 PRINT "Средняя арифметич.контроля?"
37 INPUT M1
40 PRINT "Ошибка средней контроля?"
43 INPUT O1
45 PRINT "Начат обсчет взвеш. данных"
50 LET M = 0
53 FOR I = 1 TO N
55 LET M = M + X(I)
57 NEXT I
60 LET M = M / N
63 LET S = 0
65 FOR I = 1 TO N
67 LET S = S + (M - X(I)) * (M - X(I))
70 NEXT I
73 LET D = S / (N - 1)
75 LET S = SQR(D)
77 LET O = S / SQR(N)
80 LET G = S / (SQR(2 * N))
83 LET V = (100 * S) / M
85 LET K = V * (SQR((.5 + .0001 * V * V) / N))
87 LET P = 100 * O / M
90 IF K1 = 0 GOTO 103
93 LET K3 = (M - M1) / SQR(O * O + O1 * O1)
103 LPRINT TAB(15); A$
104 LPRINT TAB(15); "N =", N
105 LPRINT TAB(15); "Ср.ар. с ошибкой", M, O
107 LPRINT TAB(15); "С к о. с ошибкой", S, G
110 LPRINT TAB(25); "Дисперсия", D
113 LPRINT TAB(15); "Кф.вар.с ошибкой", V, K
115 LPRINT TAB(15); "Пок точн.опыта", P
117 IF K1 = 0 GOTO 140
120 LPRINT TAB(15); "Критерий Стюд ", K3
140 STOP
150 END

```

Сравнение 2 рядов по Н.А. Плохинскому

```

1 PRINT "РАЗЛИЧИЯ 2 РЯДОВ"
5 PRINT "НАЗВАНИЕ 1 РЯДА"
10 INPUT AS
15 PRINT "НАЗВАНИЕ 2 РЯДА"
20 INPUT BS
25 PRINT "ДЕЙСТВУЮЩИЙ ФАКТОР"
30 INPUT FS
35 DIM A(40, 10), B(40, 10)
40 PRINT "ЧИСЛО ГРАДАЦИЙ ФАКТОРА"
45 INPUT K
50 PRINT "ЧИСЛО ПОВТОРНОСТЕЙ"
55 INPUT N
61 PRINT "ВВОД РЯДА А"
62 FOR J = 1 TO K
63 PRINT "ОЧЕРЕДНОЙ ВАРИАНТ"
64 FOR I = 1 TO N
65 INPUT X
66 A(I, J) = X
67 NEXT I
68 NEXT J
70 PRINT "ВВОД РЯДА В"
75 FOR J = 1 TO K
80 PRINT "ОЧЕРЕДНОЙ ВАРИАНТ"
85 FOR I = 1 TO N
90 INPUT Y
95 B(I, J) = Y
97 NEXT I
99 NEXT J
100 N1 = N * K * 2
105 W = N / 2
110 W1 = W * K
115 U = N1 - 2 * K
120 V2 = 0
125 FOR J = 1 TO K
130 V1 = 0
135 FOR I = 1 TO K
137 V1 = V1 + A(I, J) * A(I, J)
140 NEXT I
145 V2 = V2 + V1
150 NEXT J
155 T2 = 0
160 FOR J = 1 TO K
165 T1 = 0
170 FOR I = 1 TO N
175 T1 = T1 + B(I, J) * B(I, J)
180 NEXT I
185 T2 = T2 + T1
190 NEXT J
195 V = V2 + T2
200 H = 0
205 FOR J = 1 TO K
210 H1 = 0
215 FOR I = 1 TO N
220 H1 = H1 + A(I, J)
225 NEXT I
230 H = H + H1 * H1 / N
235 NEXT J
240 G = 0
245 FOR J = 1 TO K
250 G1 = 0
255 FOR I = 1 TO N
260 G1 = G1 + B(I, J)
265 NEXT I
270 G = G + G1 / N
275 NEXT J
280 H = H + G
285 DIM M(10)
290 FOR J = 1 TO K
295 M1 = 0
300 FOR I = 1 TO M
305 M1 = M1 + A(I, J)
310 NEXT I
315 M1 = M1 / N
320 M(J) = M1
325 NEXT J
330 DIM L(10)
335 FOR J = 1 TO K
340 L1 = 0
345 FOR I = 1 TO N
350 L1 = L1 + B(I, J)
355 NEXT I
360 L1 = L1 / N
365 L(J) = L1
370 NEXT J
375 DIM D(10), D2(10)
380 FOR J = 1 TO K
385 R = M(J) - L(J)
390 D(J) = R
395 D2(J) = R * R
400 NEXT J
405 S = 0
410 FOR J = 1 TO K
415 S = S + D(J)
420 NEXT J
425 S = S * W
430 T4 = S * S / W1
435 T5 = 0
440 FOR J = 1 TO K
445 T5 = T5 + D2(J)
450 NEXT J

```

```

455 O = T5 * W
460 C = V - H
465 C1 = C / U
470 F1 = T4 / C1
475 U1 = 1
480 U2 = U
485 F2 = (D - T4) / ((K - 1) * C1)
490 U6 = K - 1
495 U7 = U
500 PRINT "ДОСТОВЕРН.ВЛИЯНИЯ В РЯДАХ"
505 S4 = N * M1
510 S5 = N * L1
515 S6 = (S4 * S4) / N
520 S7 = (S5 * S5) / N
525 C4 = H - G - S6
530 C5 = G - S7
535 C7 = V2 - (H - G)
540 C8 = T2 - G
545 K4 = K - 1
550 N4 = N * K - K
555 Z1 = C4 / (C4 + C7)
560 Z2 = C5 / (C5 + C8)
565 F5 = (C4 / C7) * (N4 / K4)
570 F6 = (C5 / C8) * (N4 / K4)
575 LPRINT F$, A$, B$
580 LPRINT "КРИТЕРИЙ РАЗЛИЧИЯ СР.УРОВНЯ"
585 LPRINT "F1=", F1; "V1="; U1; "V2="; U2
590 LPRINT "КРИТЕРИЙ НЕПАРАЛЛ.ПРОЦЕССОВ"
595 LPRINT "F2=", F2; "V1="; U6; "V2="; U7
600 LPRINT "ДОСТОВЕРНОСТЬ ВЛИЯНИЯ ФАКТОРА"
605 LPRINT "ПОКАЗАТЕЛИ СИЛЫ ВЛИЯНИЯ"
610 LPRINT "В РЯДУ А"; Z1; "В РЯДУ В"; Z2
615 LPRINT "ДОСТОВЕРНОСТЬ ПОКАЗАТЕЛЕЙ"
620 LPRINT "В РЯДУ А"; F5; "В РЯДУ В"; F6
625 LPRINT "V1=K-1=", K4; "V2=K*N-K="; N4
630 LPRINT
635 STOP
645 END

```

Редактор Е.А.Краснова
Компьютерная верстка, макет Л.М.Кавеленова

ЛР № 020316 от 04.12.96. Подписано в печать 15.12.98.
Формат 60х84/16. Бумага офсетная. Объем 2,3 усл.изд.л., 2,5 уч.-изд.л.
Тираж 100 экз. Заказ № 122
Издательство «Самарский университет», 443011, г.Самара,
ул.Акад.Павлова, 1.
УОП СамГУ ПЛД № 67-43 от 19.02.98.