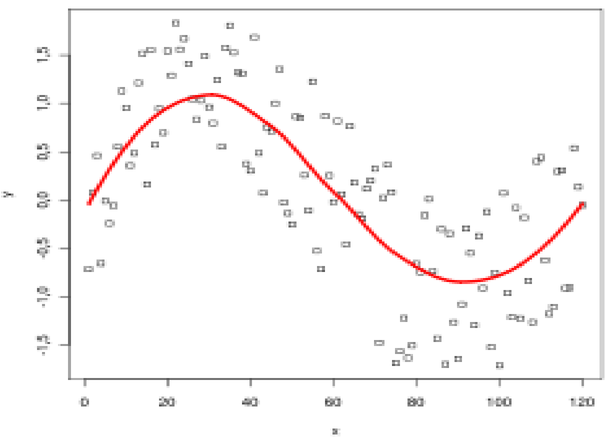
ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ В ИССЛЕДОВАНИЯХ ПО АГРОХИМИИ И АГРОПОЧВОВЕДЕНИЮ



Новосибирск 2022

Новосибирский государственный аграрный университет

Агрономический факультет

ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ В ИССЛЕДОВАНИЯХ ПО АГРОХИМИИ И АГРОПОЧВОВЕДЕНИЮ

Учебно-методическое пособие

издание второе, переработанное и дополненное

Новосибирск 2022

УДК 631.8: 519.22(07)
ББК 40.3: 22.172, я 7
О -753

Кафедра почвоведения, агрохимии и земледелия

Составитель д-р. с.- х. наук, доц. *Л.П. Галеева*
Рецензент д-р с.-х. наук, проф. *Р.Р. Галеев*

Основы математической статистики в исследованиях по агрохимии и агропочвоведению: учеб.-метод. пособие / Новосиб. гос. аграр. ун-т. Агроном. фак.; сост: Л. П. Галеева – Новосибирск: Изд-во НГАУ «Золотой колос», 2022. – 73 с.

Учебно-методическое пособие содержит теоретические и практические материалы для изучения современных методов статистической обработки экспериментальных данных в исследованиях по агрохимии и агропочвоведению.

Предназначено для проведения лабораторно-практических занятий и выполнения самостоятельной работы студентами агрономических вузов очной и заочной форм обучения по направлению подготовки Агрономия и Агрохимия и агропочвоведение.

Утверждено и рекомендовано к изданию учебно-методическим советом агрономического факультета (протокол № 2 от 30 сентября 2022 г).

ВВЕДЕНИЕ

Изучение раздела «Основы статистического анализа в исследованиях по агрохимии и агропочвоведению» дисциплины «Основы научных исследований в агрохимии и агропочвоведении» необходимо студентам по направлению подготовки Агрономия и Агрохимия и агропочвоведение для того, чтобы они были готовы проводить почвенные, агрохимические и агроэкологические исследования.

В процессе освоения данного курса студент должен освоить методы агрохимических исследований: полевой, лизиметрический, вегетационный; анализ растений, удобрений, агрохимический анализ почвы; факторы жизни растений и методы их регулирования;

1. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Математическая статистика – наука, разрабатывающая математические методы систематизации и использования статистических данных для научных и практических выводов.

Во многих своих разделах математическая статистика опирается на теорию вероятностей, которая даёт возможность оценить надёжность и точность выводов, делаемых на основании ограниченного статистического материала (например, оценить необходимый объём выборки опытных данных для получения результатов требуемой точности при выборочном обследовании).

1.1. Краткая история развития математической статистики

Ещё в **Древней Греции** высоко чтили математику, считая, что всё в природе упорядочивается в соответствии с числами.

Аристотель в своих работах уже рассуждал о корреляциях. **Леонардо да Винчи** ещё в 15 веке более конкретно высказывался об использовании математики в исследованиях. Он считал, что без применения математики недостоверна ни одна наука.

В 17 веке **Я. Бернулли** показал, что при большом числе измерений средняя арифметическая становится постоянной.

Кривая вероятности впервые была приведена в литературе **Лапласом** в 1783 г., а в 1795 г. **К. Гаусс** исследовал кривую распределения и ввёл способ наименьших квадратов.

Дальнейшая история развития математической статистики связана с трудами **Ф. Гальтона**, который в 1889 г. разработал методику **корреляционного и регрессионного анализов.** Его работы продолжил **К. Пирсон**, который развил учение о кривых распределения в биологии и предложил метод x^2, а **В. Госсет** (псевдоним **Стьюдент**) обосновал в 1908 г. теорию «малой выборки».

Особый вклад в математическую статистику внёс английский математик **Р. Фишер.** В 1935 г. он опубликовал **методику математического планирования экспериментов**, а в 1938 г. – теорию статистической проверки гипотез – **дисперсионный анализ**.

Продолжателями работ Р. Фишера стали **Ф. Йейтс**, много сделавший для разработки схем дисперсионного анализа, а также **Дж. У. Снедекор**, **Т. Литтл**, **Ф. Хиллз** и др.

В России методы статистической обработки в агрономических исследованиях впервые использовал в 1867-1869 гг. **Д.И. Менделеев**. Полную сводку методов математической статистики в 1909-1911 гг. составил **А.В. Леонтович.**

С 1929 г. эти методы пропагандировал Н.Ф. Деревицкий, он же первым изложил идеи и методы Р.Фишера. С 1931 г. распространение методов математической статистики в агрономии связано с именем **В.Н. Перегудова**.

В 60-х годах прошлого века А.А. Любищев в работе «Руководство по применению в биологии дисперсионного анализа Р. Фишера» дал глубокий анализ использования методов математической статистики. Позже появляются серьёзные работы Н.А. Плохинского (1970), П.Ф. Рокицкого (1967), В.Ю. Урбаха (1964), А.С. Молостова (1965) и др. Особое значение имел выход в 1965 г. книги Б.А. Доспехова «Методика полевого опыта», которая переиздавалась более 5 раз и до сих пор используется в исследованиях по агрономии студентами и научными работниками. Ценным пособием считается работа Г.Ф. Лакина «Биометрия» (1990).

Математическая статистика – это один из разделов математики. Она позволяет делать умозаключения о всей (генеральной) совокупности на основе наблюдений над выборочной совокупностью, или выборкой. Все статистические методы основаны на **теории вероятностей**.

Теория вероятностей – наука, изучающая общие закономерности в массовых случайных явлениях различной природы. Она применяется везде, где приходится иметь дело *с планированием экспериментов и обследований, с оценкой параметров и проверкой гипотез, с принятием решений при изучении сложных систем*. Слово «случайный» употребляется здесь для обозначения явления, исход которого в настоящий момент нельзя предсказать. Так, результаты опытов, кроме изучаемых влияний, всегда подвержены тем или иным посторонним воздействиям. Поэтому любой опыт содержит некоторый элемент случайности, который измеряется величиной экспериментальной ошибки.

Знание современных методов статистической обработки необходимо не только для количественной характеристики наблюдений и полученных в опыте данных, когда уже нельзя ничего исправить, но и на всех этапах эксперимента – от планирования до интерпретации окончательных результатов. Часто периодическое появление «модных» агротехнических приёмов, препаратов и способов быстрого повышения урожайности сельскохозяйственных культур, которые при широком применении не оправдывают возлагавшихся на них надежд, в большинстве можно объяснить отсутствием статистически обоснованных исследований.

Сами по себе методы математической статистики, если они не сочетаются с предварительным квалифицированным анализом агрономической сущности изучаемого явления и правильной постановкой опытов, не могут ничего добавить к умению экспериментатора. Никакая статистическая обработка материалов не может заставить плохой опыт дать хорошие результаты. **Главная обязанность экспериментатора** – постановка добротных, целенаправленных опытов, а ***математическая статистика помогает агрономическому исследованию в выборе оптимальных условий для проведения опыта, даёт объективную, количественную оценку экспериментальным данным.***

1.2. Основные понятия и задачи математической статистики

Основные понятия: **математика, статистика, корреляция, средняя арифметическая, кривая вероятности, способ наименьших квадратов, теория «малой выборки» и др.**

Результаты исследований в агрономии анализируют методами математической статистики, т.е. систематизируют, обрабатывают и на их основе делают обоснованные выводы и предложения. При этом используют определённые **понятия, термины и символы**.

Объекты исследований в агрономии – отдельные растения, их группы и среда произрастания. Всем им свойственно **явление изменчивости,** или **варьирования.**

Степень варьирования, выраженную математически, называют **вариацией**. Если тысячи семян одной и той же культуры, одного сорта посеять и выращивать в одинаковых условиях, то растения всегда будут отличаться по высоте, массе, внешнему виду, урожайности, её качеству и т.д.

Число таких растений или других объектов исследований представляет собой **генеральную совокупность**.

Для того, чтобы точно определить среднюю высоту растения или среднее число клубней на нём, необходимо было бы за короткий срок, за несколько часов, измерить тысячи растений и сосчитать десятки тысяч клубней, что практически невозможно. Нецелесообразно также проращивать все семена, предназначенные для посева, чтобы определить их всхожесть. В таких случаях следует воспользоваться **теорией вероятностей**, которая обобщает закономерности массовых случайных явлений. Согласно этой теории, вместо сплошного учёта всей генеральной совокупности большого объёма данных для изучения можно брать определённую её часть и судить по ней о состоянии совокупности в целом. Таким образом, по вероятности одних случайных событий находят вероятность других, связанных с первыми.

Например, в ящике 100 клубней картофеля: 30 шт. – сорта Белла и 70 шт. – сорта Лина. Какова вероятность того, что первый взятый наугад из ящика клубень будет принадлежать сорту Белла или Лина?

Вероятность взять клубень сорта Белла составит 30:100 = 0,3, а Лина – 70:100 = 0,7. Таким образом, **вероятность наступления определённого события** есть отношение чисел всех возможных случаев к общей выборке (общая выборка 100, возможные случаи – 30 и 70).

События, вероятность которых составляет более 0,5, называют ***вероятными***, а менее 0,5 – ***маловероятными***. В данном примере взять клубень сорта Лина вероятно (0,7), а Белла – маловероятно (0,3).

Отношение числа случаев с данным событием **n** к числу всех возможных случаев **N** составляет ***уровень вероятности*** **P**; $\mathbf{P = n : N}$.

Вероятность невозможного события равна 0. Например, вероятность вынуть из ящика клубни сорта Волжанин, которых там не было (**n = 0**), составит **P = 0/ N = 0.** Так, если в ящике все 100 клубней принадлежат одному сорту (n =100; N = 100), то **P = n / N** = 1. Вероятность, равная 1, называется ***достоверной***.

Нормальное распределение. Если часть (выборка) генеральной совокупности признаков (показателей) составляет менее 30 членов и стремится к бесконечности (n → ∞), то для такой части используют закономерности больших чисел, установленные для кривой нормального распределения (распределения Гаусса), показанной на рис. 1. – Кривая нормального распределения теоретического (μ, δ) и эмпирического ($\bar{x}$, S).

Площадь под кривой, находящуюся на t стандартных отклонений влево и вправо от $\bar{x}$, называют **уровнем вероятности** и выражают в %, либо в долях единицы.

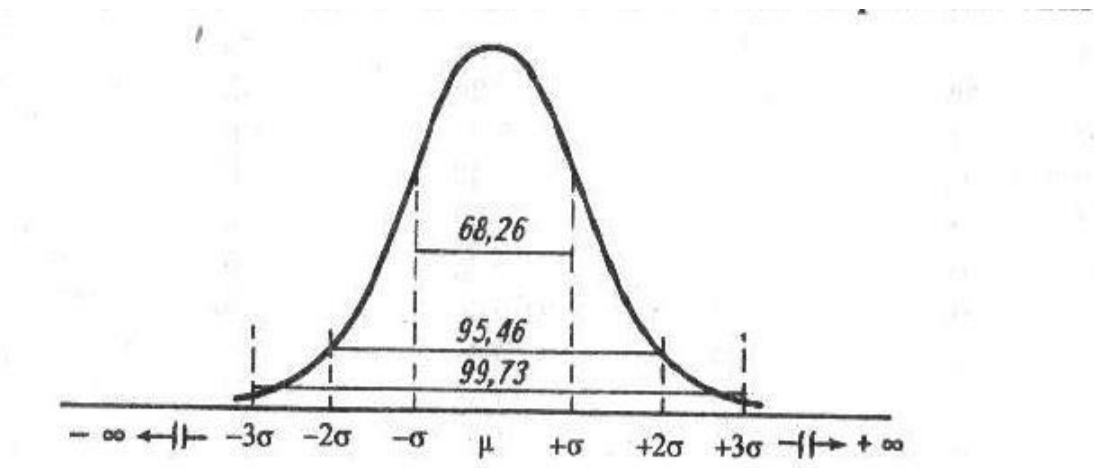


Рис.1. **Кривая нормального распределения теоретического (μ, σ) и эмпирического ($\bar{x}$, S)**

Для анализа результатов полевых опытов пользуются **уровнем доверительной вероятности**, равным 0,95, который записывают символом $P_{0,95}$, а для более ответственных анализов – уровнем 0,99 ($P_{0,99}$). На уровне доверительной вероятности $P_{0,95}$ исследователь, утверждая или отрицая какое-либо явление, положение, рискует ошибиться в 5 случаях из 100, на уровне $P_{0,99}$ – в 1 случае из 100.

Иногда пользуются не уровнем доверительной вероятности, а **уровнем значимости P_1,** который рассчитывают по формуле **P_1 = 1 - P.** Эмпирические распределения в отличие от теоретических не всегда симметричны. У них может быть не одна, а несколько вершин. Это свидетельствует о неоднородности выборки: в неё могут попасть, например, растения другого сорта или сильно отличающегося варианта.

Малые выборки. Наряду с большими выборками (**n > 30**) в агрономических исследованиях часто пользуются выборками с n < 30, например, в опыте может быть 4-8 повторностей или 10-12 вариантов (кривая нормального распределения). Выборки, состоящие менее, чем из 30 измерений и т.д., называют **малыми выборками**; на них нельзя переносить законы больших чисел. Для малых выборок применяют распределение вероятностей Стьюдента (В. Госсета), которое получило название закона малых выборок, и критерий Стьюдента, обозначаемой буквой **t**. По таблице можно найти величину вероятности, зная значение **t**. Предварительно ***рассчитывают число степеней свободы*** (**ν**) – число возможных независимых сравнений.

Например, в опыте имеется 6 повторностей (**n = 6**), от каждой из которых зависит средняя арифметическая $\bar{x}$ (**X**). Для того, чтобы получить число степеней свободы повторностей ($\boldsymbol{\nu_p}$), т.е. число не связанных один с другим элементов, уменьшают число повторностей на единицу. Тогда $\boldsymbol{\nu_p = n - 1}$. Число степеней свободы для вариантов $\boldsymbol{\nu_v}$ вычисляют по формуле $\boldsymbol{\nu_v = \ell - 1}$, где $\boldsymbol{\ell}$ - число вариантов. Для 6 повторностей опыта $\boldsymbol{\nu_p = n - 1 = 6\text{-}1 = 5}$. Критерий Стьюдента $t_{0,95} = 2{,}57$. Находим в таблице Значений t (критерия Стьюдента) число 2,57, приходим к заключению, что опыт проводится на уровне доверительной вероятности $\mathbf{P}_{\mathbf{0,95}}$, вполне достаточной для полевого опыта. Выбрав уровень вероятности и рассчитав критерий t, можно найти оптимальное число повторностей, прибавив к числу степеней свободы 1.

Пусть в опыте Р $_{0,95}$, а t $_{0,95}$ = 3,18, тогда оптимальная повторность **n** составит

$\mathbf{\nu_p + 1 = 3 + 1 = 4.}$

Критерий достоверности (существенности) – это показатель, позволяющий судить о надёжности выводов, подтверждающих или опровергающих статистическую гипотезу. Чаще всего пользуются ***нулевой гипотезой*** ($\mathbf{H_0}$). **Нулевая гипотеза** – предположение об отсутствии реального различия между фактическими и ожидаемыми (теоретическими) наблюдениями. Например, различия между средними значениями вариантов по урожайности. Его качеству, росту растений и т.д. Для 2-х средних арифметических $\mathbf{x}_1$ и $\mathbf{x_2}$ нулевую гипотезу записывают так: $\mathbf{x_1 - x_2 = 0.}$

Для проверки статистических гипотез используют **критерий достоверности.** Синонимы термина **«достоверность»** – существенность, иногда значимость, несомненность, весомость. Большинство специалистов по математической статистике рекомендуют использовать понятие **«критерий достоверности»**, некоторые – **«критерий существенности»**.

Для проверки нулевых гипотез используют **параметрические** и **непараметрические критерии.**

Параметрические критерии достоверности применимы лишь для нормального распределения, это ***критерий Стьюдента*** (t), ***критерий Фишера*** (F).

Критерий достоверности *Стьюдента* (t) прямо пропорционален разности средних арифметических $\mathbf{(x_1 – x_2)}$ или разности между долями $\mathbf{(p_1 – p_2)}$ и обратно пропорционален ошибке разности $\mathbf{(s_d).}$ Расчётное, фактическое значение критерия **t** сравнивают с теоретическими значениями на определённых уровнях значимости (табл. Значения критерия **t** на 5-ти и 1%-ном уровнях значимости).

Критерий достоверности *Фишера* (F) прямо пропорционален дисперсии вариантов ($\mathbf{S^2_v}$) и обратно пропорционален дисперсии остатка ($\mathbf{S^2_z}$). Его фактическое значение сравнивают с теоретическим, которое находят в таблице Значение критерия F на 5%-ном (вероятность 95%) и 1%-ном (вероятность 99%) уровнях значимости.

Не все выборки из биологических объектов распределяются нормально, поэтому для проверки нулевых гипотез используют **непараметрические критерии:** χ^2-критерий, **T**-критерий, критерий знаков (**Z**), которые здесь не рассматриваются.

Основные задачи статистики

Математическую статистику используют, прежде всего, для планирования опытов. В хорошо спланированном опыте должно быть достаточное число вариантов и повторностей, а все варианты в начале опыта должны находиться в одинаковых условиях. Очень важен выбор метода статистической обработки результатов.

Существенная задача математической статистики – отобрать в спланированном и заложенном опыте объекты для исследований, которые будут объективно отражать влияние изучаемых факторов. В данном случае речь идёт об использовании **метода рендомизации** при отборе образцов для опыта.

Не менее важно определить число образцов для исследований, т.е. ***оптимизировать объём выборки***.

В процессе предварительной обработки данных иногда приходится восстанавливать выпавшие данные (числа), а также браковать сомнительные. Для этого в малых выборках используют критерий тау (**τ**), а в больших – интервальную оценку средних арифметических по формуле $\bar{x}$ +- tS_x.

В проведённом опыте определяют достоверность различий между средними арифметическими исследуемых выборок. Эти задачи решают с применением критериев достоверности **t, F,** а также наименьшей существенной разности **(НСР)**. Во многих исследованиях возникает необходимость определить взаимосвязи и зависимости между различными показателями, для чего используют **коэффициент корреляции (r)** и **корреляционное отношение (η).**

Прогнозируют или отыскивают неизвестные показатели по уже известным с помощью регрессионного анализа, составляя уравнения регрессии для линейных и криволинейных зависимостей.

Почти во всех исследованиях возникает вопрос о точности опытов. Её характеризуют значением относительной ошибки $\mathbf{S_{\bar{x}}}$**%.**

Важно, что математическую статистику можно применять лишь в методически спланированных и проведённых опытах. Опыты, не отвечающие этим условиям, следует немедленно браковать.

1.3. Виды изменчивости, их характеристика и значение

Всякое массовое, множественное явление, например, группа растений на поле или животных на ферме, представляет собой совокупность особей, случаев, фактов, предметов, т.е. некоторых условных единиц, каждая из которых в отдельности строго индивидуальна и отличается от других рядом признаков – высотой, массой, количеством продукции и т.д. Каждый из признаков может иметь у различных особей разную степень выраженности, поэтому говорят, что признак варьирует. ***Свойства условных единиц*** – растений, урожаев на параллельных делянках полевого опыта и т.п. – ***отличаться друг от друга даже в однородных совокупностях называется изменчивостью,*** *или* ***варьированием***. ***Изменчивость*** – свойство, присущее всем предметам природы: двух совершенно одинаковых предметов не существует, хотя различия между ними и могут быть незаметными для невооружённого глаза.

У растений варьирующими признаками являются, например, высота, количество и масса зёрен в колосе, содержание протеина и др. Варьирование возникает вследствие того, что растения одного и того же сорта всегда отличаются своей наследственностью, а их формирование часто протекает в относительно различных условиях внешней среды. В полевых и вегетационных опытах даже при самой тщательной работе урожаи на параллельных делянках или в сосудах всегда получаются разные. Это ***колебание, изменчивость, вариация*** – результат влияния различного сочетания внешних условий, не всегда поддающихся учёту, и определяемое часто как следствие случайных причин, вызывающих различия в изучаемых признаках. Следовательно, при любом исследовании данные опытов будут всегда варьировать в тех или иных пределах.

Изменчивость, варьирование признаков создаёт известную трудность в тех случаях, когда требуется дать общую характеристику определённой варьирующей группе (совокупности) растений, животных, почв и т.п. по отдельным признакам или сравнить две такие группы и найти различие между ними. Не всегда можно (а практически очень редко) исследовать по тому или другому признаку все особи, всю совокупность. В этих случаях прибегают к изучению части её, по которой делают общее заключение. Такой метод называется ***выборочным*** и считается основным при статистическом изучении совокупности.

Таким образом, всю группу объектов, подлежащую изучению, называют ***совокупностью*** или ***генеральной совокупностью***, а ту часть объектов, которая попала на проверку, исследование, – ***выборочной совокупностью*** или просто ***выборкой***. Число элементов в генеральной совокупности и выборке называют их ***объёмом***.

Главная цель выборочного метода – по статистическим показателям малой выборки (средней) пробе возможно точнее охарактеризовать всю совокупность объектов, которая в статистике и называется генеральной совокупностью.

Цель выборочного метода научного исследования – при помощи сравнительно ограниченных средств, которые дают возможность изучать единичные явления, установить характерные свойства и законы для бесконечного числа возможных или встречающихся явлений.

В результате наблюдений мы получаем сведения о численной величине изучаемого признака у каждого члена данной выборочной совокупности. Возможные значения варьирующего признака **X** называют вариантами и обозначают $\mathbf{X_1}$, $\mathbf{X_2}$, **…**, $\mathbf{X_n}$. Полученный таким образом ряд варьирующих величин можно упорядочить – расположить значения признака (варианта) в порядке их возрастания (или убывания). Такое упорядочение ряда, т.е. расположение вариант в порядке возрастания (или убывания) называется его ***ранжированием***. После ранжирования нетрудно заметить, что каждое значение признака встречается неодинаковое число раз – одни редко, другие часто. Числа, которые характеризуют, сколько раз повторяется каждое значение признака у членов данной совокупности, называют ***частотами признака*** и обозначаются **f**. Сумма всех частот ($\sum_f$) равна объёму выборки, т.е. числу членов ряда – **n.** В результате такой обработки первичных наблюдений получаем так называемый ***вариационный ряд.***

Вариационный ряд **–** такой ряд данных, в котором указаны возможные значения варьирующего признака в порядке возрастания или убывания и соответствующие им частоты.

Различают два типа изменчивости: ***количественную***, которая может быть измерена, и ***качественную***, которая не поддаётся измерению.

Количественная изменчивость – такая, в которой различия между вариантами выражаются количеством. Количественно могут изменяться масса урожая, процентное содержание сахара, кислот, витаминов, крахмала или белка в урожае, размеры растений, содержание питательных элементов в почве, т.е. всё, что имеет массу, размер, объём и т.п. Различают 2 вида количественной изменчивости: ***прерывистую***, или ***дискретную*** и ***непрерывную***.

При ***прерывистой (дискретной)*** изменчивости различия между вариантами выражаются целыми числами, между которыми нет и не может быть переходов, например, число растений на квадратном метре, число зёрен в колосе и т.д. Примером ***дискретной метрической изменчивости*** служат: количество побегов, корней, почек, усов, цветков, кистей, продуктивных стеблей, плодов в штуках на растение, побег, ветку, кисть, колос или $м^2$; засорённость посевов, численность вредителей и густота стеблестоя в $шт/м^2$.

При ***непрерывной изменчивости*** значения вариант выражаются мерами объёма, длины, массы и т.д., между которыми мыслимы любые переходы с неограниченным числом возможных значений; всё зависит от степени точности, принимаемой для характеристики данного количественного признака.

Ход анализа вариационных рядов количественной изменчивости зависит от объёма выборки – малого (менее 30 единиц) или большого (более 30). Как для малых, так и для больших выборок вычисляют следующие основные статистические характеристики: среднюю арифметическую $\bar{x}$, дисперсию $\mathbf{S^2}$, стандартное отклонение **S**, ошибку средней арифметической $\mathbf{S}_{\bar{x}}$, коэффициент вариации **V**, относительную ошибку средней арифметической $\mathbf{S}_{\bar{x}}\%$. В конце анализа дают интервальную оценку средней арифметической.

Малые выборки

Примером малых выборок может быть число повторностей, которое чаще всего колеблется от 3 до 6. К малым выборкам относится также число колосков в колосе, клубней картофеля в кусте, семян гороха в бобах и т.п.

Для малых выборок вычисляют такие статистические характеристики: средние арифметические, дисперсии, стандартные отклонения, коэффициенты вариации, ошибки выборочных средних, относительные ошибки и др.

Средняя арифметическая простая ($\bar{x}$). Для её вычисления используем, например, данные учёта числа листьев кукурузы на полях с рядковым и разбросным внесением удобрений (табл. 1).

Таблица 1. Зависимость числа листьев кукурузы на одном растении от способа внесения удобрений, шт

Внесение удобрений (вариант)	Повторение				Средняя арифметическая $\bar{x}$
	1	2	3	4	
Рядковое	14	18	13	15	15
Разбросное	16	7	20	17	15

Варьирующий показатель – число листьев – обозначают буквой **X**, количество повторений – **n**. Среднюю арифметическую простую ($\bar{x}$) вычисляют по формуле $\bar{x}$ **= ∑ X: n.**

Для варианта с рядковым внесением удобрений средняя арифметическая $\bar{x}_1$ = ∑ X/ n = (14+18+13+15): 4 = 15 листьев.

Для варианта с разбросным внесением удобрений $\bar{x}_2$ = ∑ X: n = (16+7+20+17): 4 = 15 листьев.

Итак, средние арифметические количества листьев кукурузы при разных способах внесения удобрений одинаковые – 15, но ***размах вариации*** **(R)** разный. При рядковом внесении удобрений он составит $\mathbf{R_1}$ **= X max – X min = 18-13 = 5**, а при разбросном внесении удобрений $\mathbf{R_2}$ **= X max – X min =**

20-7 = 13. Следовательно, число листьев варьирует более значительно при разбросном внесении удобрений.

Средняя арифметическая является основной статистической характеристикой вариационного ряда, все остальные лишь объясняют её.

Дисперсия. При повторных исследованиях для других выборок одной и той же совокупности размах вариации может быть другим, т.е. он не является характерным показателем варьирования. Более полно вариационный ряд характеризует ***дисперсия*** $\mathbf{S^2}$ – средний квадрат отклонений каждого члена вариационного ряда $(\mathbf{x_1,\ x_2,\ \ldots,\ x_n})$ от средней арифметической $\bar{x}$. Дисперсию вычисляют по формуле $\mathbf{S^2 = \sum (X - \bar{x})^2 : (n - 1)}$.

Для вычисления дисперсии составляют таблицу (табл. 2).

Таблица 2. Вычисление квадрата отклонений от х

Рядковое внесение удобрений				Разбросное внесение удобрений			
Повторение	X_1	$X-\bar{x}$	$(X-\bar{x})^2$	Повторение	X_2	$X-\bar{x}$	$(X-\bar{x})^2$
1	14	-1	1	1	16	1	1
2	18	3	9	2	7	-8	64
3	13	-2	4	3	20	5	25
4	15	0	0	4	17	2	4
	15	0	14		15	0	94

Подставив данные в формулу, получаем дисперсии для каждого варианта (способа внесения удобрений):

$\mathbf{S_1}^2$ = 14: (4-1) = 4,67 листа и $\mathbf{S_2}^2$ = 94: (4-1) = 31,3 листа.

Дисперсии характеризуют не только величину изменения вариационных рядов, но и специфику варьирования. Дисперсия для 2-го варианта почти в 7 раз больше дисперсии для 1-го (31,3 : 4,67 ≈7).

Стандартное отклонение (S). Этот показатель представляет собой корень квадратный из дисперсии и вычисляется по формуле $\boldsymbol{S = \sqrt{S^2}}$. В вариантах с рядковым внесением удобрений $\mathbf{S_1} = \sqrt{4{,}67} = 2{,}16$, с разбросным внесением $\mathbf{S_2} = \sqrt{31{,}3} = 5{,}59$. Как и дисперсия, стандартное отклонение выражается в тех же единицах измерения, что и характеризуемый им признак. Чем сильнее варьирует показатель, тем больше числовое значение стандартного отклонения. В расчётах оно является более удобной характеристикой, чем дисперсия.

Коэффициент вариации. Дисперсия и стандартное отклонение непригодны, когда в опытах сравнивают изменчивость признаков, выраженных разными единицами измерения (масса урожая, т; число плодов, шт.; длина побегов, см; площадь листьев, $см^2$ и др.). В этих случаях пользуются ***коэффициентом вариации***. ***Коэффициент вариации V*** – это отношение стандартного отклонения **S** к средней арифметической $\bar{x}$, выраженное в процентах: $\mathbf{V = 100 \cdot (S : \bar{x})}$. Коэффициент вариации 1-го варианта составляет $\mathbf{V_1 = 100 \cdot (2{,}16 : 15) = 14{,}4\%}$, 2-го – $\mathbf{V_2 = 100 \cdot (5{,}59 : 15) = 37{,}3\%}$, т.е. при разбросном внесении удобрений варьирование более чем в 2 раза выше, чем при рядковом (37,3:14,4 ≈ 2,6).

Сравним вариационный ряд урожайности, где средняя арифметическая $\bar{x}_1$ = 1,06 т, а стандартное отклонение $\mathbf{S_1}$ = 0,18 т, с вариационным рядом площади листьев, где средняя арифметическая $\bar{x}_2$ = 32,4 $см^2$, а стандартное отклонение $\mathbf{S_2}$ = 2,5 $см^2$. Коэффициент вариации урожайности составит $\mathbf{V_1 = 100 \cdot (0{,}18 : 1{,}06) = 17\%}$, а площади листьев – $\mathbf{V_2 = 100 \cdot (2{,}5 : 32{,}4) = 7{,}7\%}$.

Если судить только по стандартным отклонениям, то можно сделать ошибочный вывод о том, что урожайность варьирует меньше, чем площадь листьев. Однако коэффициенты вариации свидетельствуют о том, что урожайность варьирует в 2 раза сильнее, чем площадь листьев.

Коэффициент вариации не превышает 50%, если распределение симметрично, при сильно ассиметричных распределениях он может достигать 100% и более. ***Варьирование незначительное***, если коэффициент вариации находится в ***пределах 10%***, ***среднее***, если он равен ***10-20%,*** и ***значительное***, если он ***превышает 20%***. Коэффициент вариации используют для расчёта объёма выборки, числа повторностей при планировании опытов.

Ошибка выборочной средней **($S_{\bar{x}}$).** Средние арифметические имеют ошибки, которые возникают в результате неполной представительности выборки. Эти ошибки свойственны только выборочному методу исследования, их числовое значение зависит как от степени изменчивости изучаемого признака, так и от объёма выборки. Ошибку выборочной средней вычисляют по формуле

$$S_{\bar{x}} = \sqrt{\frac{\sum (X - \bar{x})^2}{n \cdot (n-1)}}.$$

В нашем примере с внесением удобрений (2 варианта) и 4 повторностями (n = 4), где уже вычислены значения $\sum (X-\bar{x})^2$, ошибка выборочной средней для 1-го варианта (рядковое внесение удобрений) составит

$S_{\bar{x}1} = \sqrt{14 / 4 (4-1)} = 14/12 = 1{,}08$,

а для 2-го варианта (разбросное внесение удобрений)

$S_{\bar{x}2} = \sqrt{94 / 4 (4-1)} = 94/12 = 2{,}8$.

Ошибку стандартного отклонения $\mathbf{S_{\bar{x}s}}$ рассчитывают по формуле $\mathbf{S_{\bar{x}s} = S : \sqrt{2n}}$. Формула для расчёта ошибки коэффициента вариации имеет вид $\mathbf{S_{\bar{x}v} = V : \sqrt{2n}}$.

Все эти ошибки выражают в тех же единицах, что и варьирующий признак, приписывая их к соответствующим характеристикам $\mathbf{x \pm S_{\bar{x}}}$.

Таким образом, определяют предельное числовое значение средней арифметической. Если это значение хотят определить с вероятностью 0,99, то $\mathbf{S_{\bar{x}}}$ утраивают (увеличивают в 3 раза): $x \pm 3\mathbf{S_{\bar{x}}}$. Тогда на уровне вероятности $P_{0,99}$ наименьшее значение средней арифметической для 1-го варианта будет $x - 3\mathbf{S_{\bar{x}}} = 15 - 3 \cdot 1{,}08 = 11{,}76$ листа, а наибольшее $x + 3\,\mathbf{S_{\bar{x}}} = 15 + 3 \cdot 1{,}08 = 18{,}24$ листа. На уровне вероятности $P_{0,95}$ предельное значение определяют, вычитая или прибавляя к средней арифметической $2\mathbf{S_{\bar{x}}}$, тогда $x_1 - 2\mathbf{S_{\bar{x}}} = 15 - 2 \cdot 1{,}08 = 12{,}84$ листа, а $x_2 + 2\mathbf{S_{\bar{x}}} = 15 + 2 \cdot 1{,}08 = 17{,}16$ листа.

Относительная ошибка выборочной средней $S_{\bar{x}}$ % – это отношение ошибки выборочной средней к соответствующей средней арифметической, выраженное в %: $S_{\bar{x}}$ % = 100 · ($S_{\bar{x}}$ √ x).

Чем меньше относительная ошибка, тем выше точность средней арифметической. Точность принято считать **высокой** при $S_{\bar{x}}$ % **≤3%**, **средней** при $S_{\bar{x}}$ % **= 3-6%** и **низкой** – при $S_{\bar{x}}$ % > **6-7%.**

Сомнительными считаются результаты полевых опытов, в которых значение относительных ошибок выражается десятыми и сотыми долями процентов. Это может говорить либо о погрешности в расчётах, либо, к сожалению, о недобросовестности исследователя.

Большие выборки

Для больших выборок статистические характеристики можно вычислять способом произведений. Но этот способ трудоёмок, особенно если числа многозначны, а объём выборки очень большой. В таком случае более удобен ***способ условной средней***, т.е. от произвольного начала **А**. При этом способе все данные ранжируют, выделяют группы с определённым интервалом **i,** определяют частоту **f,** т.е. количество членов в каждой группе вариационного ряда.

Вариационный ряд – ряд ранжированных чисел, для которых указаны значения варьирующего признака и соответствующие им частоты (т.е. сколько раз повторяется тот или иной признак).

Пример. В опыте с яровой пшеницей взяли подряд 40 колосьев (n = 40), измерили их длину (в см) и результаты разместили в возрастающем порядке: **4,2;** 4,5; 4,6; 4,7; 4,9; 5,4; 5,5; 5,6; 5,8; 5,9; 6,0; 6,0; 6,3; 6,4; 6,5; 6,7; 6,7; 6,8; 6,9; 6,9; 7,1; 7,2; 7,2; 7,2; 7,3; 7,3; 7,4; 7,4; 7,8; 7,8; 7,9; 8,2; 8,5; 8,7; 8,8; 9,1; 9,6; 9,9; 10,1; **10,7**.

Обработку вариационного ряда ведут в такой последовательности:

1. Определяют число групп по формуле **Ч**$_г$ = √ **n** = √ 40 ≈ 6-7 групп. Как правило, когда **n** находится в пределах 30-60, берут 6-7 групп, 60-100 – 7-8 групп, более 100 – 8-15 групп.

2. Вычисляют интервал групп (колосьев) по формуле

i = ($\bar{x}_{max}$ **–** $\bar{x}_{min}$**) : Ч**$_г$ = (10,7 - 4,2) : 6 ≈ 1 см.

Значения интервала групп для удобства расчётов округляют до целого числа. Последующие расчёты проводят, записывая результаты в таблицу 3.

Таблица 3. Таблица для обработки данных вариационного ряда (большие выборки)

№ группы	Интервал группы, см	Средние значения группы, $\bar{x}$	Частота, **f**	Отклонение, $\bar{x}$**-A**	**f (**$\bar{x}$**-A)**	**(**$\bar{x}$**-A)**^{2}	**f (**$\bar{x}$**-A)**^{2}
1.	4,2-5,2	4,7	5	-2,2	-11,0	4,84	24,20
2.	5,3-6,3	5,8	8	-1,1	-8,8	1,21	9,68
3.	6,4-7,4	**6,9 (A)**	15	0	0	0	0
4.	7,5-8,5	8,0	5	1,1	5,5	1 ,21	6,05
5.	8,6-9,6	9,1	4	2,2	8,8	4,84	19,36
6.	9,7-10,7	10,2	3	3,3	9,9	10,89	32,67
			∑f = 40		∑f **(**$\bar{x}$**-A)** **= 4,4**		∑f **(**$\bar{x}$**-A)**^{2} **= 91,96**

Первая группа начинается наименьшим числом (4,2), к которому прибавляют интервал **i,** составляющий 1 см., последующие группы образуют аналогично. Вычисляют среднее значение каждой группы ($\bar{x}$), а одно из них, которое чаще всего встречается в выборке (**f**), берётся за произвольное начало (**A**). Оно должно быть в середине групп и иметь наибольшую частоту. Таким числом будет 6,9, его частота (**f**) составляет 15.

Сумма всех чисел должна быть равна объёму выборки **(n)** **∑**$_{\mathbf{f}}$ **= n** = 40. Вычитая значения из произвольного начала, получают отклонения $\bar{x}$**-A** (со знаком + или -).

Дальнейшие расчёты ведут по формулам:

- **произвольный момент первой степени $\mathbf{b} = \sum_f (\bar{x}-A) : n$** = 4,4 : 40 = +0,11;

- **средняя арифметическая $\bar{x} = A + b$** = 6,9 +0,11 = 7,01 см;

- **корректирующий фактор $C = [\sum_f (\bar{x}-A)]^2 : n$** = $4{,}4^2$ / 40 = 0,484;

- **дисперсия $S^2 = \sum_f (\bar{x}-A)^2 : (n - 1)$** = 91,96 – 0,484 / (40-1) = 2,35;

- **стандартное отклонение $S = \sqrt{S^2}$** = √ 2,35 = 1,53 см;

- **коэффициент вариации $V = \sqrt{100S} : \bar{x}$** = 100 · 1,53 /7,01= 21,8%;

- **ошибка выборочной средней $S_{\bar{x}} = S : \sqrt{n}$** = 1,53 : √ 40 = 1,53 : 6,325 = 0,242 см;

- **относительная ошибка среднего арифметического $S_{\bar{x}}\ \% = (S_{\bar{x}} \cdot 100) : \bar{x}$** = 0,242·100 : 7,01 = 3,45%;

Интервальную оценку средней арифметической дают на двух уровнях вероятности: $\mathbf{P_{0,95}}$ и $\mathbf{P_{0,99}}$.

Вычисляют число степеней свободы $\boldsymbol{\nu = n - 1}$ = 40-1= 39. По таблице Стьюдента (Приложение 1) находят теоретические значения критерия $\mathbf{t_{0,95}}$ и $\mathbf{t_{0,99}}$, которые подставляют в формулу для интервальной оценки $\bar{x} \pm t \cdot S_{\bar{x}}$. Произведение $t \cdot S_{\bar{x}}$ называют ***областью индивидуального рассеивания***. Подставив в формулу вычисленные значения, а также $\mathbf{t_{0,95}}$ = 2,04 и $\mathbf{t_{0,99}}$ =2,75, имеем интервал для $\mathbf{P_{0,95}}$ 6,52-7,50 см и $\mathbf{P_{0,99}}$ 6,34-7,65 см. Изображают вариационный ряд графически (рис. 2) и делают **выводы:**

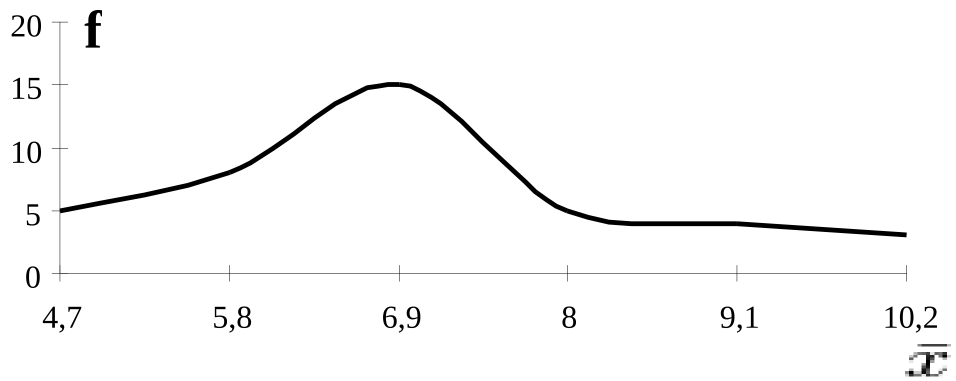


Рис. 2. Кривая вариационного ряда выборки признака (длина колоса)

1) **средняя арифметическая ($\bar{x}$) длины колоса** равна 7,01 см;

2) **коэффициент вариации V** = 21,8% – свидетельствует о значительной вариации длины колосьев;

3) **значение относительной ошибки среднего арифметического $S_{\bar{x}}$**% = 3,45 указывает на то, что средняя арифметическая вычислена с удовлетворительной точностью;

4) **к данному вариационному ряду** относятся колосья с длиной в интервале 6,52-7,50 см при уровне $\mathbf{P_{0,95}}$ и 6,34-7,65 см при уровне $\mathbf{P_{0,99}}$;

5) **на графике кривая вариационного ряда** имеет одну вершину, что свидетельствует об однородности выборки.

Качественная изменчивость

При качественной изменчивости в выборке имеется одна из 2-х возможностей (**альтернатив**) – данный признак либо есть, либо отсутствует. Такую изменчивость называют ещё **альтернативной.** В опытах с качественной изменчивостью вместо измерения какого-либо показателя подсчитывают число объектов с тем или иным признаком. Примеры качественной изменчивости: число повреждённых или здоровых растений, число подмёрзших растений, число испортившихся и здоровых клубней картофеля в хранилище и т.п.

Качественная изменчивость – варьирование, когда различия между вариантами выражаются качественными показателями, которые одни варианты имеют, а другие нет (цвет, вкус, форма изучаемого объекта). Если признак принимает только 2 взаимоисключающих друг друга значения (больной – здоровый, остистый – безостый), то изменчивость называется ***альтернативной***, т.е. ***двояковозможной***.

Для анализа качественной изменчивости вычисляют следующие статистические характеристики:

1) долю наличия признака **p**;
2) долю отсутствия признака **q;**
3) показатель изменчивости качественного признака **S;**
4) коэффициент вариации признака $\mathbf{V_p}$;
5) ошибку выборочной доли $\mathbf{S_p}$.

Общий объём выборки обозначают буквой **N**, а число объектов с данным признаком – **n.**

Доля наличия признака – это отношение числа объектов с данным признаком (**n**) к общему числу объектов (**N**), т.е. к объёму выборки, рассчитанное по формуле **p = n : N.**

Например, в выборке из 100 клубней картофеля сорта Лина 30 шт. поражены паршой, а в выборке из 100 клубней картофеля Белла – 10 шт. Обозначив характеристики для сорта Лина $\mathbf{p_1, n_1, N_1}$, а для сорта Белла $\mathbf{p_2, n_2, N_2}$, получим

$\mathbf{p_1 = n_1 : N_1} = 30 : 100 = 0{,}3$; $\mathbf{p_2 = n_2 : N_2} = 10 : 100 = 0{,}1$.

Доля отсутствия признака – это разность между целым, т.е. 1 и долей наличия признака, рассчитанная по формуле $\mathbf{q = 1 - p}$. Для сорта Лиина $\mathbf{q_1 = 1 - p_1}$ = 1- 0,3 = 0,7; для сорта Белла $\mathbf{q_2 = 1 - p_2}$ = 1- 0,1 = 0,9.

Показатель изменчивости качественного признака для альтернативной изменчивости, т.е. когда изучаемый объект имеет две градации, как в нашем случае (поражённые и непоражённые клубни), рассчитывают по формуле $\mathbf{S = \sqrt{p \cdot q}}$.

Для сорта Лина $\mathbf{S_1 = \sqrt{p_1 \cdot q_1}} = \sqrt{0{,}3 \cdot 0{,}7} = 0{,}458$, а для сорта Белла $\mathbf{S_2 = \sqrt{p_2 \cdot q_2}} = \sqrt{0{,}1 \cdot 0{,}9} = 0{,}3$.

Максимальная изменчивость наблюдается при $\mathbf{p = q = 0{,}5}$. При этом показатель **S** изменчивости также равен **0,5** $\mathbf{S_{max}} = \sqrt{0{,}5 \cdot 0{,}5} = 0{,}5$.

Если изучаемый объект имеет более 2-х градаций, например, в выборке есть плоды томата зелёные, бурые, спелые и перезревшие, то показатель изменчивости вычисляют по формуле

$$\mathbf{S = \sqrt[k]{p_1 \cdot p_2 \cdot \ldots \cdot p_k}},$$

где $\mathbf{p_1}$, $\mathbf{p_2}$, ... $\mathbf{p_k}$ – доли признака от общего объёма выборки; **k** – число градаций признака.

Так, если из 100 плодов томата 15 оказалось зелёных, 35 бурых, 40 спелых и 10 – перезревших, то при k = 4, $S = \sqrt[4]{15 \cdot 35 \cdot 40 \cdot 10} = 0{,}214$.

Коэффициент вариации $\mathbf{(V_p)}$ **–** отношение показателя изменчивости **(S)** к его максимальному значению $\mathbf{(S_{max})}$, выраженное в %, вычисляют его по формуле $\mathbf{V_p = 100 \cdot (S : S_{max})}$.

Для сорта Лина коэффициент вариации составит $\mathbf{V_{p1}} = 100 \cdot (0{,}458:0{,}4) = 91{,}6\%$, а для сорта Белла $\mathbf{V_{p2}} = 100 \cdot (0{,}3:0{,}5) = 60\%$. Максимальное значение коэффициента вариации 100% наблюдается при $\mathbf{S = S_{max}} = 0{,}5$.

Ошибка выборочной доли **(Sp) –** это мера отклонения от доли наличия признака, её для альтернативной изменчивости вычисляют по формуле

$Sp = \sqrt{}\ p \cdot q : N$.

Для изучаемых сортов картофеля ошибка выборочной доли составит **$Sp_1 = \sqrt{}\ p_1 \cdot q_1 : N_1$** $= \sqrt{}\ 0{,}3 \cdot 0{,}7 : 100 = 0{,}046$; **$Sp_2 = \sqrt{}\ p_2 \cdot q_2 : N_2$** $= \sqrt{}\ 1 \cdot 0{,}9 : 100 = 0{,}03$.

Интервальную оценку доли дают по формуле **$p \pm 2Sp$** на уровне доверительной вероятности **$P_{0,95}$** и **$p \pm 3Sp$** на уровне вероятности **$P_{0,99}$**.

Эти интервалы для сорта Лина при **$P_{0,95}$** составят $0{,}3 - 2 \cdot 0{,}046 = 0{,}3 - 0{,}092 = 0{,}208$; $0{,}3 + 2 \cdot 0{,}046 = 0{,}392$ доли.

Итак, нижняя граница интервала составила 0,208 доли, а верхняя – 0,392 доли. Если градаций более 2-х, то ошибку выборочной доли вычисляют по формуле **$Sp = S : \sqrt{}\ N$,** где **S** – показатель изменчивости, а **N** – объём выборки.

Для примера со 100 плодами томата, имеющими 4 градации зрелости и **S** = 0,214, значение ошибки выборочной доли составит **$Sp = S : \sqrt{}\ N$** $= 0{,}214 : \sqrt{}\ 100 = 0{,}214 : 10 = 0{,}0214$.

В пределах двойной ошибки выборочной доли все значения доли укладываются с вероятностью $P_{0,95}$, в пределах тройной ошибки – с вероятностью $P_{0,99}$.

1.4. Подготовка данных к статистической обработке

Перед статистической обработкой данные необходимо соответствующим образом подготовить:

1) округлить,

2) вычислить средние арифметические по каждой делянке и варианту,

3) выбраковать сомнительные данные,

4) восстановить выпавшие данные,

5) преобразовать данные.

1. Округление опытных данных. В исследованиях пользуются следующим правилом: для получения достаточно точных чисел, необходимо иметь опытные данные с 3-мя значащими цифрами. Так, урожайность следует записывать 0,187; 1,87 и 18,7 т/га.

Для более тщательного округления используют уменьшенное в 4 раза стандартное отклонение определённого вариационного ряда. Если первой значащей цифрой для **S:4** окажется целое число, то данные округляют до целого числа.

При расчёте суммы квадратов берут дополнительную цифру, т.е. если исходные данные имеют десятые доли, то квадраты вычисляют до сотых. Если цифра за последней значащей цифрой больше 5 или после 5 следует цифра больше 0, то последнюю значащую цифру увеличивают на 1. Так, числа 84,67 и 84,651 округляют до 84,7. Если за последней значащей цифрой стоит 5, а затем нули, то последнюю значащую цифру увеличивают на 1: 84,550 = 84,6, а чётная цифра остаётся неизменной: 84,450 =84,4.

2. Вычисление средних арифметических. Вычисление простых средних арифметических было показано ранее. Иногда в опытах урожай собирают с разных площадей. Так, на участке площадью 0,5 га урожайность картофеля ($\bar{x}_1$) составила 13,0 т/га, а на участке площадью 16 га ($\bar{x}_2$) – 11 т/га. ***Средняя арифметическая***, рассчитанная по формуле простой $\bar{x}$ = (13,0 + 11,0):2 =12,0 т/га. Но так как площади участков очень различаются, то следует вычислять ***взвешенную среднюю арифметическую***

$$\bar{x}_{\text{взв.}} = \frac{x_1f_1 + x_2f_2 + \ldots + x_nf_n}{\sum f} = \frac{13,0 \cdot 0,5 + 11,0 \cdot 16}{0,5+16} = \frac{6,5 + 176,0}{16,5} = 11,1 \text{ т/га.}$$

Таким образом, она будет на 0,9 т меньше простой средней арифметической (12,0 - 11,1 = 0,9). После вычисления средних арифметических по каждой опытной делянке необходимо проверить гипотезу о принадлежности сомнительных данных к совокупности. Речь идёт о браковке тех данных, которые достоверно отличаются от всех остальных в конкретных вариационных рядах.

3. Браковка сомнительных данных. Сомнительные данные, которые значительно отличаются от всех остальных данных какого-либо варианта, определяют только с помощью математической статистики. Субъективная браковка данных недопустима. Рассмотрим это на конкретном примере.

Например, в вегетационном опыте, где была 6-кратная повторность в варианте с двойной дозой азота, учли массу растений и получили следующие результаты, г/сосуд: 20,8; 19,0; 10,1; 19,9; 21,0; 22,0. Чтобы убедиться, что все данные принадлежат к определённому вариационному ряду, выполняют следующие операции:

1) ранжируют вариационный ряд массы растений в возрастающем порядке: 10,1; 19,0; 19,9; 20,8; 21,0; 22,0.

Двум первым и двум последним данным присваивают обозначения X_1, X_2, X_{n-1}, X_n. Наиболее сомнительными будут крайние, т.е. 10,1 и 22,0. Их сомнительность проверяют путём расчёта **τ (тау)** для X_1 и X_n по формулам

$$\boldsymbol{\tau_1} = \frac{X_2 - X_1}{X_{n-1} - X_1} = \frac{19{,}0 - 10{,}1}{21{,}0 - 10{,}1} = 0{,}817;$$

$$\boldsymbol{\tau_n} = \frac{X_n - X_{n-1}}{X_n - X_2} = \frac{22{,}0 - 21{,}0}{22{,}0 - 19{,}0} = 0{,}333;$$

2) находят критерии **τ** теоретические (Приложение 4). При значении **n** = 6 $\boldsymbol{\tau_{0,95}} = 0{,}689$, $\boldsymbol{\tau_{0,99}} = 0{,}805$;

3) сравнивают критерии **τ** расчётные с теоретическими. Если **τ** расчётные больше или равны **τ** теоретическим, то проверяемые данные сомнительны.

ВЫВОДЫ:

1) так как $\boldsymbol{\tau_1} = 0{,}817$, т.е. больше $\boldsymbol{\tau_{0,95}} = 0{,}689$ и $\boldsymbol{\tau_{0,99}} = 0{,}805$, то проверяемый результат (10,1) сомнителен и должен быть выбракован. Это утверждается не только на уровне доверительной вероятности $\mathbf{P_{0,95}}$, но и на уровне $\mathbf{P_{0,99}}$, когда по теории вероятностей возможность ошибиться составляет 1 случай на 100;

2) τ_n = 0,333, т.е. меньше $\tau_{0,95}$ = 0,689 и $\tau_{0,99}$ = 0,805, следовательно, проверяемые данные 22,0 не сомнительны и не должны браковаться.

Итак, среднюю арифметическую следует вычислять не из 6, а из 5-ти оставшихся данных: (19,0+19,0+20,8+21,0+22,0) : 5 = 20,7 г.

Без браковки сомнительных данных средняя арифметическая была бы значительно заниженной: $\bar{x}$ = (10,1+19,0+19,0+20,8+21,0+22,0) : 6 = 18,8 г.

Следует отметить, что браковать сомнительные данные по приведённой методике, можно при количестве повторностей не менее 4.

4. Восстановление выпавших данных. Обработка данных часто осложняется выпадением данных на некоторых делянках опыта (повреждения растений птицами, вредителями, затопление делянки после ливней, проезд транспорта и т.д.).

Из-за выпадения данных средние арифметические в вариантах могут быть либо завышены, если выпал результат с очень низким числовым значением, либо занижены, если выпал результат с наибольшим числовым значением. Из-за этого возникают ошибки, которые можно устранить, если восстановить выпавшие данные, т.е. вычислить их наиболее вероятные значения. При выпадении одного результата пользуются формулой

$$\bar{x}_{\text{вос.}} = \frac{\ell V + nP - \sum X}{(\ell - 1)(n - 1)},$$

где ℓ - число вариантов;

V – сумма данных в варианте, где выпал результат;

n – число повторностей в опыте;

P – сумма данных в повторностях, где выпал результат;

$\sum X$ – сумма данных во всём опыте, за исключением выпавшего результата.

Для примера возьмём урожайность зерна кукурузы и результаты запишем в таблицу 4. Для удобства расчётов в этом и последующих примерах будем выражать урожайность в ц/га.

Таблица 4. Зависимость урожайности кукурузы от внесения разных доз азота, ц/га

Вариант	Урожайность по повторностям		
	1	2	3
1. Контроль (без удобрений)	40	39	40
2. N 30	39	X	42
3. N 60	42	44	43
4. N 90	43	47	46

Как видно, во 2-й повторности 2-го варианта отсутствует результат (X), и его надо восстановить.

V = 39+42 =81;

P = 39+44+47 = 130;

$\sum X$ = 465, тогда

$$\bar{x}_{\text{вос.}} = \frac{(4 \cdot 81) + (3 \cdot 130) - 465}{(4-1)\cdot(3-1)} = 41{,}2 \text{ ц/га.}$$

Восстановленный результат ставят на выпавшее место и выполняют соответствующий анализ. При выпадении одновременно нескольких данных в одном опыте можно использовать метод статистической обработки для опытов с неполным числом данных.

5. Преобразование исходных данных. Не все результаты исследований подчиняются **законам нормального распределения**, иногда имеется неоднородность дисперсий по выборкам, наблюдается большое варьирование по вариантам опыта. В таких случаях проводят следующие преобразования.

Если данные в опытах, где учитывают, например, численность сорняков, вредителей, распространение болезней, выражаются большими числами, их преобразуют путём извлечения корня квадратного их X ($\sqrt{X}$). Например, если в почве в одном из повторений число семян сорняков составляет 4231 шт/м^2, тогда $\mathbf{X_{преобр.} = \sqrt{4231} = 65}$.

Если результаты исследований выражены числами, близкими к нулю, их преобразуют по формуле $\mathbf{\sqrt{X} + 1}$.

Если некоторые наблюдения равны 0, например, баллом 0 выражают отсутствие подмерзания растений, результаты таких наблюдений трансформируют в lg X или lg (X + 1).

Если наблюдаемые показатели выражены в %, например, степень поражения листьев болезнью, результаты учётов преобразуют в угол-арксинус √ % (Приложение 5). Такие преобразования проводят при числовых значениях показателей от 0 до 15 и от 85 до 100%, так как здесь можно сильно снизить варьирование. Данные в пределах 15-85% в преобразовании не нуждаются. Так, если поражение листьев болезнью составило 9,3%, то угол-арксинус √ % будет равен 17,8.

Данные опытов, в которых определяют действие повреждающих факторов, например, летальные дозы радиации, преобразуют в пробиты. Так, при гибели 67% растений пробит составляет 5,44.

Преобразованные данные обрабатывают методом дисперсионного анализа, оценивают достоверность разности между вариантами, а затем возвращаются к исходным данным.

1.5. Выбор метода статистической обработки данных

Если данные по каждой экспериментальной единице, например, по опытной делянке, не кажутся сомнительными, вычисляют средние арифметические по вариантам и проводят агрономический анализ всего опыта. Если ничто в опыте не вызывает сомнений, его результаты подвергают соответствующему анализу, например, дисперсионному.

Метод анализа результатов полевых опытов зависит от способа размещения вариантов. При размещении вариантов методом рендомизированных повторений используют дисперсионный анализ рендомизированных повторений, при полной рендомизации – дисперсионный анализ полной рендомизации, при размещении латинским квадратом – дисперсионный анализ латинского квадрата и т.д.

Многофакторные опыты можно размещать методами рендомизированных повторений, расщеплённых делянок, смешивания. Результаты подвергают соответствующим дисперсионным анализам – рендомизированных повторений, расщеплённых делянок, смешивания.

Данные вегетационных опытов, а также опытов по технологии хранения продукции обрабатывают теми же методами, что и данные полевых опытов, размещённых методом полной рендомизации. Результаты опытов, в которых варианты размещены стандартными методами, обрабатывают **разностным методом**, предназначенным для сопряжённых выборок.

Если варианты в опыте размещены систематически (что, вообще-то не рекомендуется), результаты обрабатывают дробным методом. При обработке показателей качественной изменчивости достоверность разницы определяют по критерию Стьюдента.

Соответствие между наблюдаемыми и теоретическими распределениями в генетических исследованиях оценивают по критерию χ^2.

Зависимости между различными показателями растений, между растениями и средой их произрастания определяют с помощью **корреляционных** и **регрессионных анализов.** Для изучения корреляции качественных признаков используют формулу **Юлла.**

Дисперсионный анализ – метод анализа результатов эксперимента, заключающийся в разложении общей изменчивости результативного признака, например, урожайности, на части – компоненты, соответствующие повторениям, вариантам, ошибкам случайного порядка и т.д. Значимость действия и взаимодействия изучаемых факторов оценивают по **F** и **НСР** $_{05}$.

Предположим, что проведён полевой опыт, размещённый методом рендомизированных повторений. Основной его показатель – урожайность – изменяется по вариантам, повторениям, а также из-за случайных причин – неучтённого варьирования условий среды и самих растений.

Р. Фишер выразил эти изменения суммами квадратов следующих рассеиваний: вариантов – **Cv**; повторений – **Cр**; ошибки – **Cz**. Их сложение даёт сумму квадратов общего рассеивания: **Cy = Cv + Cр + Cz**.

Для каждого рассеивания вычисляют количество (число) степеней свободы **ν** по формулам:

$$\nu_y = N - 1; \quad \nu_v = (\ell - 1); \quad \nu_p = (n - 1); \quad \nu_z = (\ell - 1)(n - 1).$$

Разделив суммы квадратов на соответствующее число степеней свободы, получают дисперсию S^2. «Дисперсия» означает рассеивание данных опыта и расчленение общего варьирования изучаемых показателей на составные части. Отсюда и название метода – **дисперсионный анализ**.

Для дисперсионного анализа представляют интерес дисперсия вариантов S_v^2 и дисперсия ошибки S_z^2. Соотношение дисперсий – это тот критерий, который позволяет дать общую оценку достоверности различий между средними арифметическими опыта. В честь автора дисперсионного анализа критерий обозначили первой буквой его фамилии **F** (критерий Фишера). **F** вычисляют по формуле $F = S_v^2 : S_z^2$.

Расчётный фактический критерий $F_{факт.}$ сравнивают с теоретическим $F_{теор.}$, который находят по таблицам (Приложение 2 и 3). Если $F_{факт.} \geq F_{теор.}$, то достоверность различий в опыте доказана, т.е. имеется одна или несколько пар вариантов, средние арифметические которых достоверно различаются. Если $F_{факт.} < F_{теор.}$, то достоверных различий между вариантами нет.

Бывают случаи, когда $\mathbf{F}_{\text{факт.}}$ лишь несколько меньше $\mathbf{F}_{\text{теор.}}$ Строго следуя правилу, изложенному выше, можно сделать вывод об отсутствии достоверной разницы в опыте. Однако дальнейший анализ позволяет иногда найти эту разницу хоты бы в одном-двух вариантах. В таких случаях, не ограничиваясь расчётом **F,** следует продолжить вычисления до наименьшей существенной разности (НСР). С этим показателем сравнивают разницу между парами вариантов (d $x_1 - x_2$). Если d ≥ НСР, то разницу между анализируемыми вариантами считают доказанной. Доказательства чаще всего ведут на уровнях доверительной вероятности $\mathbf{P}_{0,95}$ и $\mathbf{P}_{0,99}$.

Дисперсионный анализ – наиболее совершенный метод статистической обработки данных, но он применим только к опытам, размещённым методом рендомизации.

Преимущества дисперсионного анализа:

1) вычленение из общего варьирования его компонентов;

2) вычисление обобщённой ошибки всего опыта **Е** на основе большого числа наблюдений, чем индивидуальные ошибки отдельных вариантов в недисперсионных методах.

Например, при 5 вариантах и 4-х повторностях опыта число степеней свободы ошибки составит (5-1)·(4-1) = 12, тогда как для каждого варианта опыта 4-1 = 3.

Дисперсионный анализ особенно ценен для многофакторных опытов, так как позволяет определить достоверность не только действия, но и взаимодействия факторов.

В конце дисперсионного анализа обязательно вычисляют относительную ошибку всего опыта $\mathbf{S}_{\bar{x}}\,\mathbf{\%} = \mathbf{100\ (E : \bar{x}_N)}$,

где Е – обобщённая ошибка опыта, $\bar{x}_N$ – средняя арифметическая всего опыта.

По числовому значению относительной ошибки судят о точности опыта. Определение точности обязательно при анализе результатов любых исследований.

Основное различие дисперсионных анализов состоит в перечне вычисляемых сумм квадратов.

1) для рендомизированных повторений вычисляют $C_y = C_p + C_v + C_z$;

2) для латинского квадрата и латинского прямоугольника

$C_y = C_p + C_c + C_v + C_z$;

3) для 2-х факторного опыта $C_y = C_p + C_A + C_B + C_{AB} + C_z$;

4) для 3-х факторного опыта $C_y = C_p + C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC} + C_z$;

5) для метода смешивания $C_y = C_{p1} + C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_z$;

6) для двойного расщепления делянок $C_y = C_p + C_A + C_B + C_{AB} + C_{z1} + C_{z2}$;

7) для тройного расщепления делянок $C_y = C_p + C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC} + C_{z1} + C_{z2} + C_{z3}$.

2. АНАЛИЗ Д ОДНОФАКТОРНЫХ ОПЫТОВ

1. Опыт, размещённый методом рендомизированных повторений с полным набором данных.

Возьмём результаты урожайности кукурузы в опыте с гербицидом агелон, в котором восстановлены данные. Количество вариантов $\ell = 4$, количество повторений $n = 3$, общее количество делянок $N = \ell \cdot n = 4 \cdot 3 = 12$.

Данные опыта размещаем в таблице с учётом вариантов и повторений и находим среднюю урожайность кукурузы в вариантах как простую средннюю арифметическую величину по каждому повторению отдельно (табл. 5).

Таблица 5. Зависимость урожайности кукурузы от дозы гербицида, ц/га

Вариант, ℓ	Повторения, n			Сумма	Среднее $\bar{x}$
	1	2	3		
1. Контроль (без гербицида)	40	39	40	119	39,7
2. Гербицид, 1 кг/га	39	41	42	122	40,7
3. Гербицид, 1,5 кг/га	42	44	43	129	43,0
4. Гербицид, 2 кг/га	43	47	46	136	45,3

Среднюю арифметическую ($\bar{x}$) по всему опыту округляют до целого числа (39,7+40,7+43,0+45,3 = 506:4 = 42,17 ≈ 42), берут за произвольное начало (А = 42) и вычисляют отклонения каждого результата от произвольного начала: 40-42= -2; 39-42= -3 и т.д. Составляют таблицу отклонений по повторениям (**Р**) и вариантам (**V**) и всего опыта (**q**). Все отклонения возводят в квадрат и составляют таблицу квадратов.

Таблица 6. Отклонения результатов опыта от произвольного начала **А**

Вариант, ℓ	Повторения, n			Отклонения суммы
	1	2	3	
1	-2	-3	-2	-7
2	-3	-1	0	-4
3	0	2	1	3
4	1	5	4	10
Сумма по повторениям (Р)	-4	3	3	q = 2

Таблица 7. Квадраты отклонений и их суммы

Вариант, ℓ	Повторения, n			$\sum\alpha^2 S$	S^2
	1	2	3		
1	4	9	4	17	49
2	9	1	0	10	16

3	0	4	1	5	9
4	1	25	16	42	100
$\sum\alpha^2 p$	14	39	21	$\sum\sum\alpha^2 = 74$	$\sum S^2 = 174$
$\sum p^2$	16	9	9	$\sum p^2 = 34$	$q^2 = 4$

Средняя арифметическая всего опыта

$\bar{x}_N = \sum x : N = (40+39+40+39+41+42+42+44+43+43+47+46) : 12 = 42{,}2$ ц/га.

Суммы квадратов рассеиваний общего (C_y), повторений (Cp), вариантов (Cv) и ошибки (C_z) вычисляют по формулам:

$C_y = (\sum\sum\alpha^2 \cdot N - q^2) : N = (74 \cdot 12 - 4) : 12 = 73{,}7;$

$Cp = (\sum p^2 \cdot n - q^2) : N = (34 \cdot 3 - 4) : 12 = 8{,}2;$

$Cv = (\sum s^2 \cdot \ell - q^2) : N = (174 \cdot 4 - 4) : 12 = 57{,}7;$

$C_z = C_y - Cp - Cv = 73{,}7 - 8{,}2 - 57{,}7 = 7{,}8.$

Вычисляют число степеней свободы общего рассеивания (V_y), повторений (V_p), вариантов (V_v), остатка (V_z):

$V_y = N - 1 = 12\text{-}1 = 11$; $V_p = n - 1 = 3\text{-}1 = 2$; $V_v = \ell - 1 = 4\text{-}1 = 3$;

$V_z = (\ell - 1)\cdot(n - 1) = 3 \cdot 2 = 6.$

После этого данные заносят в таблицу дисперсионного анализа, в которой вычисляют дисперсию вариантов (S^2_v), дисперсию остатка (S^2_z) и критерий $F_{\text{факт.}}$ (расчётный).

Дисперсию рассчитывают по формулам:

$S^2_v = Cv : V_v = 57{,}7 : 3 = 19{,}2;$

$s^2_z = C_z : V_z = 7{,}8 : 6 = 1{,}3.$

Критерий Фишера фактический ($F_{\text{факт.}}$) рассчитывают по формуле

$F_{\text{факт.}} = S^2_v : S^2_z = 19{,}2 : 1{,}3 = 14{,}8.$

Результаты дисперсионного анализа записывают в таблицу (табл. 8).

Таблица 8. Результаты дисперсионного анализа

Рассеивание	Суммы квадратов	V	S^2	$F_{факт.}$	$F_{теор.}$	
					$P_{0,95}$	$P_{0,99}$
Общее C_y	73,7	11	-			
Повторений Ср	8,2	2	-	14,8	4,76	9,78
Вариантов Cv	57,7	3	19,2			
Остатка (ошибки) C_z	7,8	6	1,3			

Теоретическое значение критерия Фишера ($F_{теор.}$) находят по таблице (Приложение 2 и 3) по числу степеней свободы вариантов (V_v) – 3 (колонка с числом 3) и остатка (V_z) – 6 (шестая строчка). На их пересечении находим $F_{0,95}$ = 4,76 и $F_{0,99}$ = 9,78. Составляют итоговую таблицу дисперсионного анализа.

Сравнивая фактический (расчётный) критерий Фишера с теоретическим значением, делают вывод о достоверности опыта.

ВЫВОД: так как фактическое значение критерия Фишера составляет 14,8, что больше $F_{теор.\ 0,95}$ = 4,76 и $F_{теор.\ 0,99}$ = 9,78 (см. табл. 8), значит опыт достоверный на обоих уровнях доверительной вероятности - $P_{0,95}$ и $P_{0,99}$.

Это значит, что разности между средними арифметическими пар вариантов будут достоверными и дисперсионный анализ надо продолжать. Если критерий Фишера расчётный намного меньше теоретического значения, тогда все расчёты прекращают и делают вывод об отсутствии достоверных разностей между какими-либо парами вариантов опыта.

Обобщённую ошибку опыта (E) и ошибку разности (S_d) рассчитывают по формулам:

$$E = \sqrt{S^2_z : n} = \sqrt{1,3 : 3} = 0,66;$$

$$S_d = E \cdot 1,41 = 0,66 \cdot 1,41 = 0,93.$$

1,41 – это постоянное число, $\sqrt{2} = 1,41$.

Наименьшую существенную разность (НСР) рассчитывают, как правило, на двух уровнях доверительной вероятности по формулам:

$$НСР_{0,95} = S_d \cdot t_{0,95};\ НСР_{0,99} = S_d \cdot t_{0,99}.$$

Теоретическое значение критерия Стьюдента находят в таблице (Приложение 1) по числу степеней свободы остаточного рассеивания V_z, которое в нашем опыте составляет 6. В графе Приложения 1 выбирают уровень доверительной вероятности ($P_{0,95}$ или $P_{0,99}$), а в строке (в приведённом примере) число 6. На их пересечении находят $t_{0,95} = 2,45$, а $t_{0,99} = 3,71$.

Рассчитывают $НСР_{0,95} = S_d \cdot t_{0,95} = 0,93 \cdot 2,45 = 2,28$ ц/га; $НСР_{0,99} = S_d \cdot t_{0,99} = 0,93 \cdot 3,71 = 3,45$ ц/га.

Относительную ошибку всего опыта вычисляют по формуле

$S\bar{x}\% = 100 \cdot E : \bar{x}_N = 100 \cdot 0,66 : 42,2 = 1,56\%$.

Затем в итоговую таблицу дисперсионного анализа вносят значения средних арифметических по вариантам, $НСР_{0,95}$; $НСР_{0,99}$ и относительную ошибку опыта ($S\bar{x}\%$).

Таблица 9. Итоговая таблица дисперсионного анализа

Вариант	$\bar{x}$	Разность	НСР		$S\bar{x}\%$
			0,95	0,99	
1	39,7	-			
2	40,7	1,0	2,28	3,45	1,56
3	43,0	3,3			
4	45,3	5,6			

Сравнивая разности между парами средних арифметических со значениями НСР, делают выводы о существенности этих разностей. При этом придерживаются правила: если разность между какими-либо парами средних арифметических больше, чем значения НСР, или равна им, тогда эти разности существенны.

В данном случае (см. табл. 9) существенными будут различия по урожайности кукурузы в варианте 3 при уровне достоверности $P_{0,95}$, где разность составила 3,3 т/га, что превышает 2,28, а в варианте 4 – при обоих уровнях достоверности $P_{0,95}$ и $P_{0,99}$, так как разность в этих вариантах составила 5,6 ц/га, что превышает 2,28 и 3,45.

ВЫВОДЫ:

1) поскольку во 2-м варианте прибавка урожайности составляет 1 ц/га, что меньше $НСР_{0,95}$ (2,28 ц/га) и $НСР_{0,99}$ (3,45 ц/га), то она **недостоверна;**

2) в 3-м варианте прибавка урожая составила 3,3 ц/га, что больше $НСР_{0,95}$ (2,28 ц/га), значит на этом уровне вероятности она **достоверна**, а на высшем уровне $НСР_{0,99}$ (3,45 ц/га) – **недостоверна**.

3) в 4-м варианте прибавка урожая составила 5,6 ц/га, что больше $НСР_{0,95}$ (2,28 ц/га) и $НСР_{0,99}$ (3,45 ц/га), значит она достоверна на обоих уровнях доверительной вероятности.

Таким образом, эффективными дозами гербицида оказались 1,5 и 2 кг/га;

4) разность между 3-м и 4-м вариантами составляет 2,3 ц/га, что больше $НСР_{0,95}$ (2,28 ц/га), следовательно, целесообразность повышения дозы гербицида до 2 кг/га доказана.

Значение относительной ошибки $S_{\bar{x}}\%$, которая составляет 1,56%, свидетельствует об очень высокой точности опыта. Этими выводами и заканчивается дисперсионный анализ.

3. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Между варьирующими явлениями, объектами, условиями среды, ростом, продуктивностью растений и другими показателями существуют определённые взаимосвязи: значение средней величины одного признака изменяется при изменении другого признака. Когда определённому значению независимой переменной **Х** соответствует несколько значений другого признака **У**, зависимость приобретает стохастический характер.

Взаимосвязи между варьирующими признаками называют **корреляцией**.

Классификация корреляций. Корреляции разделяют по **направлению, форме** и **числу** связей.

По направлению корреляции: 1) **прямая**;

2) **обратная**.

При **прямой корреляции** с увеличением значения **Х** увеличивается значение признака **У**. Примеры прямой корреляции:

1) чем быстрее нарастает число клубней картофеля определённых размеров, тем выше урожайность;

2) чем больше длина листа, тем больше его площадь;

3) чем лучше освещены растения, тем интенсивнее синтез органических веществ и т.п.

При **обратной корреляции** с увеличением значения признака **Х** значение признака **У** уменьшается. Например, при постоянном увеличении массы корней свёклы уменьшается их сахаристость.

По **форме** корреляция бывает: **1) линейная** и **2) криволинейная**.

Линейная корреляция имеет место, когда с увеличением признака **Х** соответственно увеличивается второй признак **У**. Например,

1) площадь листьев возрастает с увеличением их длины;

2) урожайность увеличивается с увеличением числа полноценных зёрен;

3) ростовые процессы улучшаются при увеличении площади питания растений.

При **криволинейной корреляции** значения **Х** и **У** изменяются сначала в одном направлении, а затем в противоположных. Так, при постоянно возрастающих дозах фактора **Х** (азотные или другие удобрения, влажность почвы, её плотность и т.п.) урожайность **У** сначала возрастает, затем стабилизируется, а после дальнейшего увеличения признак X снижается (рис. 3).

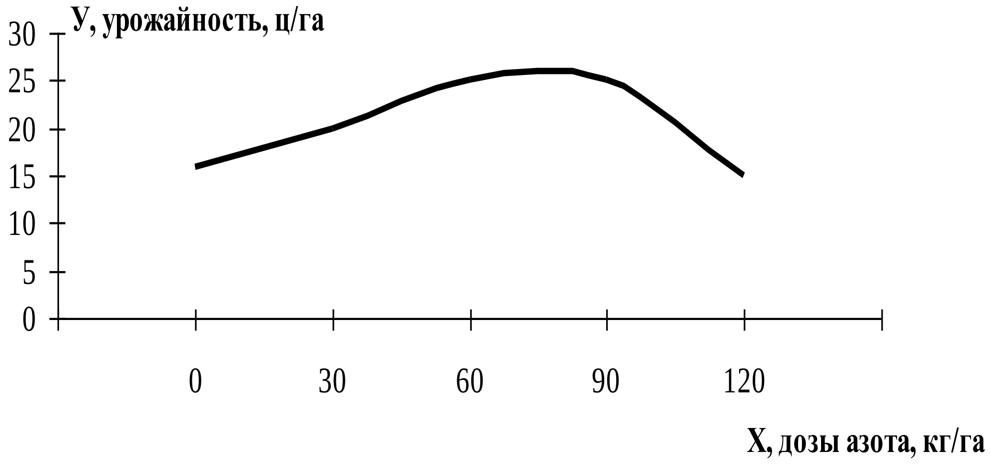


Рис. 3. Влияние доз азота на урожайность яровой пшеницы

Линейная связь выражается **коэффициентом корреляции *r***, а **криволинейная** – **корреляционным отношением *ή*** (эта).

По **числу связей** корреляция может быть **простой**, когда имеется связь между двумя признаками и **множественной,** когда связано три признака и более. Например, урожайность зависит от дозы азота, фосфора, калия, норм орошения и других факторов.

По **силе связи** корреляция бывает **полной, сильней, средней, слабой**; она также может быть **достоверной** и **недостоверной**.

Значение корреляций и регрессий. Если корреляционный анализ показал наличие сильной и достоверной связи, т.е. такой, которая установлена на уровнях вероятности $P_{0,95}$ и $P_{0,99}$, тогда проводят регрессионные анализы, вычисляя коэффициент регрессии $\mathbf{R_{xy}}$ или $\mathbf{R_{yx}}$**.**

Регрессия – это характер и степень изменения одного из признаков **X** на единицу измерения другого **У**. Например, с изменением длины листа на 1 см его площадь изменяется на 4,6 см^2.

После корреляционных и регрессионных анализов составляют ***уравнения регрессии***, которые используют:

1) для вычисления неизвестного показателя по известному, например, площади листьев по их длине;

2) для прогнозирования будущей урожайности по числу цветков или завязей;

3) для прогнозирования качества урожая по элементам погоды;

4) для прогнозирования распространения вредителей и болезней по внешним признакам;

5) для прогнозирования качества продуктов переработки и их хранения по качеству сырья и т.д.

Анализ линейной зависимости

Корреляционный анализ. Проанализируем зависимость между длиной листьев пшеницы и их площадью (табл. 10).

Таблица 10. Зависимость между длиной листьев пшеницы и их площадью

Номер пар	Длина листьев (Х), см	Площадь листьев (У), см^2	Х-$\bar{x}$	У- ȳ	(Х-$\bar{x}$)(У- ȳ)	(Х-$\bar{x}$)^2	(У- ȳ)^2
1	15,0	6,21	-6,4	-7,89	49,9	38,4	60,8
2	16,2	7,50	-5,2	-6,5	33,8	27,0	42,3
3	17,5	9,10	-3,9	-4,9	19,1	15,2	24,0
4	18,9	10,0	-2,5	-4,0	10,0	6,25	16,0
5	20,2	11,7	-1,2	-2,3	2,76	1,44	5,29
6	20,5	12,0	-0,9	-2,0	1,8	0,81	4,0
7	20,7	12,5	-0,1	-1,5	0,15	0,01	2,25
8	20,9	12,9	-0,5	-1,1	0,55	0,25	1,21
9	21,3	13,1	-0,1	-0,9	0,09	0,01	0,81
10	21,7	13,6	0,3	-0,4	-1,2	0,09	0,16
11	22,0	14,0	0,6	0,0	0	0,36	0,0
12	22,2	15,0	0,8	1,0	0,8	0,64	1,0
13	22,2	15,5	0,8	1,5	1,2	0,64	2,25
14	22,6	15,8	1,2	1,8	2,16	1,44	3,24
15	22,9	16,2	1,5	2,2	3,3	2,25	4,84
16	23,0	17,0	1,6	3,0	4,8	2,56	9,0
17	24,1	18,1	2,7	4,1	11,1	7,29	16,81
18	24,9	19,1	3,5	5,1	17,9	12,3	26,0
19	25,4	20,2	4,0	6,2	24,8	16,0	38,4
20	25,3	21,1	3,9	7,1	27,7	15,21	50,4
	$\bar{x}$= 21,4	ȳ = 14,0	∑=0	∑=0	∑= (Х-$\bar{x}$)(У- ȳ)=210	∑(Х-$\bar{x}$)^2= 148	∑ (У- ȳ)^2 = 309
Число пар n = 20							

Вычисляют:

коэффициент корреляции

$$r = \frac{\sum(X-\bar{x})(У-\bar{y})}{\sqrt{\sum(X-\bar{x})^2 \cdot \sum(У-\bar{y})^2}} = \frac{210}{\sqrt{148 \cdot 309}} = +0{,}98;$$

ошибку коэффициента корреляции

$$S_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-0{,}98^2}{20-2}} = 0{,}047;$$

критерий достоверности коэффициента корреляции

$$t_r = r : S_r = 0{,}98 : 0{,}047 = 20{,}9$$

Теоретическое значение критерия Стьюдента находят по числу степеней свободы $v_r = n - 2 = 20\text{-}2 = 18$, тогда $t_{0,95} = 2{,}1$; $t_{0,99} = 2{,}88$.

О силе связей делают вывод согласно такому правилу:

если коэффициент корреляции равен 1 (r = 1) – **связь полная**;

если он составляет 0,66-0,99 (r = 0,66-0,99) – **сильная**,

0,33-0,66 (r = 0,33-0,66) – **средняя**;

менее 0,33 (r < 0,33) – **слабая.**

О направлении связи вывод делают в зависимости от знака коэффициента корреляции: **«плюс» (+) - корреляция прямая**, **«минус» (-) - обратная.**

Вывод о достоверности связей делают на основе такого правила:

- если критерий существенности коэффициента корреляции фактический ($t_{r\ факт.}$) больше критерия теоретического ($t_{r\ теор.}$) или равен ему, то связь **достоверная** (существенная) – $t_{r\ факт.} \geq t_{r\ теор.}$

ВЫВОДЫ:

1) поскольку r = 0,98, то между длиной листьев пшеницы и их площадью **связь сильная, почти полная**;

2) коэффициент корреляции имеет положительный знак, поэтому **корреляция прямая**;

3) поскольку критерий Стьюдента фактический (t_r) составляет 20,9, что больше $t_{0,95} = 2,1$ и $t_{0,99} = 2,88$, то **связь между длиной листьев пшеницы и их площадью существенна на самых высоких уровнях доверительной вероятности.**

Если число пар незначительное, то оценка достоверности корреляции искажается. Р. Фишер предложил оценивать достоверность по критерию корреляции t_z, пользуясь формулой $\mathbf{t_z = Z \sqrt{n-1}}$. Например, n =7, а r =0,69. При этом Z = 0,848, а $t_z = 0,848 \sqrt{7-3} = 1,7$.

Число степеней свободы $v_r = n – 2 = 7 – 2 = 5$. Для $v_r = 5$ $t_{0,95} = 2,57$ и $t_{0,99} = 4,03$. Так как $t_z = 1,7$, что меньше $t_{0,95}$ и $t_{0,99}$, то **связь недостоверная**.

Для оптимизации числа пар (повторностей) при корреляционном анализе пользуются формулой

$$n_{опт.} = (t^2 : Z^2) + 3,$$

где t – критерий Стьюдента для V_r (в данном примере n – 2= 7 – 2 = 5); при этом $t_{0,95} = 2,57$, а $t_{0,99} = 4,03$;

Z – показатель, предложенный Р. Фишером (в этом примере 0,848), тогда

$n_{0,95} = 2,57^2 : 0,848^2 + 3 = 12,2 \approx 13$ пар;

$n_{0,99} = 4,03^2 : 0,848^2 + 3 = 25,6 \approx 26$ пар.

Таким образом, для проведения корреляционного анализа на уровне $P_{0,95}$ необходима выборка из 13 пар, а на уровне $P_{0,99}$ – из 26 пар.

4. РЕГРЕССИОННЫЙ АНАЛИЗ

При сильной и достоверной связи и любом её направлении (прямом или обратном) проводят регрессионный анализ, вычисляя **коэффициент регрессии R_{yx}.** Для нашего примера логично вычислить изменения площади листьев пшеницы на единицу изменения длины по формуле

$$\mathbf{R_{yx}} = \frac{\sum(X-\bar{x}) \cdot \sum(У-\bar{y})}{\sum(X-\bar{x})^2} = \frac{210}{148} = 1{,}42 \text{ см}^2 \text{ на 1 см длины.}$$

Площадь листьев (X) по их длине (Y) рассчитывают по уравнению линейной регрессии

$$\mathbf{Y = y + R_{yx} (X - \bar{x})} = 14 + 1{,}42\,(X - 2{,}14).$$

Значения **X** получают после измерения длины 20-30 листьев пшеницы и определения их средней длины. Например, среднее значение длины листа составляет 21,7 см.

Фактическое значение площади листа при этой длине составляет 13,6 $см^2$, а расчётное будет $\mathbf{Y = 14 + 1{,}42\,(21{,}7 - 21{,}4) = 14 + 0{,}43 = 14{,}4\ см^2}$**.** Разница между расчётной и фактической площадью составляет 14,4 – 13,6 = 0,8 $см^2$, или 5,9%. Таким образом, по уравнению регрессии площадь листа вычислена с удовлетворительной точностью.

Умножив площадь одного листа на их количество на растении, получим общую поверхность листьев на одном растении. Зная количество растений на определённом участке, можно подсчитать на нём площадь всех листьев.

5. АНАЛИЗ КРИВОЛИНЕЙНОЙ ЗАВИСИМОСТИ

В некоторых случаях, например, при изучении влияния норм высева семян на величину урожайности, наблюдается такая закономерность: с увеличением нормы высева урожайность растёт, при какой-то определённой норме он стабилизируется, а при дальнейшем увеличении нормы начинает снижаться из-за загущенности посевов. Подобная связь называется ***криволинейной.*** Если её анализировать с помощью коэффициента корреляции **r**, он может указать на отсутствие связи или наличие весьма слабой зависимости. Для криволинейной зависимости вычисляют **корреляционные отношения** $\acute{\eta}_{xy}$ и $\acute{\eta}_{yx}$**.**

Вычисление корреляционного отношения. Например, в опыте с горохом изучали нормы высева и урожайность зелёных бобов. В таблице 11 приведены вспомогательные данные для расчёта корреляционного отношения по формуле

$$\acute{\eta}_{yx} = \frac{\sqrt{(У-\bar{y})^2 - (У-\bar{y}_x)^2}}{(У-\bar{y})^2}$$

Таблица 11. Вспомогательные данные для расчёта корреляции между нормой высева гороха и урожайностью зелёных бобов

Номер пар	Х, Норма высева	У, Урожай-ность	$\bar{y}_x$	$(У-\bar{y}_x)$	$(У-\bar{y}_x)^2$	$У-\bar{y}$	$(У-\bar{y})^2$
1	0,8	44	50	-6	36	-27	729
2	1,0	56		6	36	-15	225
3	1,2	62	68	-6	36	-9	81
4	1,4	74		6	36	3	9
5	1,6	88	91	-3	9	17	289
6	1,8	94		3	9	23	529
7	2,0	91	85	6	36	20	400
8	2,2	79		-6	36	08	64
9	2,4	69	61	8	64	-02	4
10	2,6	53		-8	64	-18	324
			$\bar{y}$= 71	$\sum(У-\bar{y}_x) = 0$	$\sum(У-\bar{y}_x)^2=362$	$\sum(У-\bar{y}) = 0$	$\sum(У-\bar{y})^2$ =2654

Нормы высева как значения независимой переменной **Х** располагают в возрастающем порядке. Вариационный ряд разбивают на 4-7 групп так, чтобы в каждой из них было не менее 2-х наблюдений. Число наблюдений в группах может быть разное. 10 норм высева целесообразно разбить на 5 групп, как показано в таблице 11.

Вычисления:

корреляционное отношение

$$\acute{\eta}_{yx} = \frac{\sqrt{(У-\bar{y})^2 - (У-\bar{y}_x)^2}}{(У-\bar{y})^2} = \frac{\sqrt{2654 - 362}}{2654} = \mathbf{0{,}929};$$

ошибка корреляционного отношения

$$s_{\acute{\eta}} = \frac{\sqrt{1-\acute{\eta}^2 yx}}{n-2} = \frac{\sqrt{1-0{,}929^2}}{10-2} = \mathbf{0{,}131};$$

критерий достоверности корреляционного отношения фактический

$$t_{ή} = \frac{ή_{yx}}{s_{ή}} = \frac{0{,}929}{0{,}131} = \mathbf{7{,}09}.$$

Значение $t_{\text{теор.}}$ берут из табл. 1 приложений для степени свободы $v_{ή}$ = n-1 = 10-2 =8; $t_{0,95}$ = 2,31; $t_{0,99}$= 3,36.

ВЫВОДЫ:

1) значение $ή_{yx}$ = **0,929** свидетельствует о сильной связи между нормами высева семян гороха и урожайностью зелёных бобов;

2) критерий ή = 7,09 больше $t_{0,95}$ = 2,31 и $t_{0,99}$ = 3,36, следовательно, **связь достоверна на обоих уровнях вероятности**.

Корреляционное отношение всегда выражается положительным числом, поэтому выводы о направлении связи не делают.

Коэффициент корреляции **r** для данного примера составляет лишь 0,23, что значительно меньше $ή_{yx}$, следовательно, для криволинейных зависимостей необходимо вычислять только корреляционное отношение.

Составление уравнений регрессии для криволинейной зависимости

Поскольку в примере с урожайностью гороха связь сильная и высокодостоверная, используем эти данные для составления уравнения регрессии. Графическое изображение такой зависимости имеет форму параболы и выражается квадратическим уравнением

$$\mathbf{\text{У} = \frac{\text{у} \sum(X-\bar{x})\cdot\text{У}\;\;(X-\bar{x})}{\sum(X-\bar{x})^2} + \frac{[\;\sum(X-\bar{x})^2\text{У} - n\,C\text{У}\;]\;[(X-\bar{x})^2 - C]}{\sum(X-\bar{x})^4 - n\,C^2}}$$

где $\mathbf{C = \sum(X-\bar{x})^2 : n.}$

Из таблицы 12 берут значения средних сумм, повторностей и подставляют в формулу, предварительно рассчитав **C**:

C = 3,3 : 10 = 0,33.

После решения получаем уравнение регрессии

$$У = 71 + \frac{31,4\,(X - 1,7)}{3,3} + \frac{191,7 - 10 \cdot 0,33 \cdot 71}{1,93 - 10 \cdot 0,33^2}\,\mathbf{[(X-1,7)^2 - 0,33]} =$$

$$= 181,7\,X - 50\,X^2 - 74,9. \quad \mathbf{У = 181,7\,X - 50\,X^2 - 74,9.}$$

Таблица 12. Исходные данные для составления уравнения квадратичной параболы

Норма высева гороха, ц/га (X)	Урожайность зелёных бобов гороха, ц/га (У)	$X-\bar{x}$	$(X-\bar{x})^2$	$(X-\bar{x})^4$	$(X-\bar{x})У$	$(X-\bar{x})^2У$
0,8	44	-0,9	0,81	0,6561	-39,6	35,64
1,0	56	-0,7	0,49	0,2401	-39,2	27,44
1,2	62	-0,5	0,25	0,0625	-31,0	15,5
1,4	74	-0,3	0,09	0,0081	-22,2	6,66
1,6	88	-0,1	0,01	0,0001	-8,8	0,88
1,8	94	0,1	0,01	0,0001	9,4	0,94
2,0	91	0,3	0,09	0,0081	27,3	8,19
2,2	79	0,5	0,25	0,0625	39,5	19,75
2,4	69	0,7	0,49	0,2401	48,3	33,81
2,6	53	0,9	0,81	0,6561	47,7	42,93
$\bar{x}$ = 1,7	$\bar{y}$ = 71	$\sum$=0	$\sum(X-\bar{x})^2$ = 3,3	$\sum(X-\bar{x})^4$ = 1,93	$\sum(X-\bar{x})У$ =31,4	$\sum(X-\bar{x})^2У$ = 191,7

Для проверки вычисления урожайности зелёных бобов гороха по этому уравнению возьмём норму высева 2 ц/га. Подставив эту норму в уравнение, получим расчётную урожайность

$У = 181,7 \cdot 2 - 50,65 \cdot 2^2 - 74,9 = 363,4 - 202,6 - 74,9 = 85,9$ ц/га. Фактическая урожайность 91 ц/га. Ошибка прогнозирования составляет 91-85,9 = 5,1 ц/га, или 5,6%. Точность прогнозирования урожайности **удовлетворительная**.

Кроме кривых типа параболы могут быть кривые типа гиперболы, логарифмические и кривые многих других типов. Если простое уравнение кривой подобрать не очень легко, то проводят выравнивание ряда **У** способом **простой скользящей средней**.

6. МНОЖЕСТВЕННАЯ КОРРЕЛЯЦИЯ

Если результат опыта зависит одновременно от нескольких показателей или факторов, то имеет место **множественная корреляция**.

Например, зависимость урожайности зерна кукурузы **У** от массы початков молочно-восковой спелости **X**, массы листьев **Z.** Силу связи между ними определяют по коэффициенту множественной корреляции **R** на основании коэффициентов корреляции для парных связей r_{yx}, r_{yz}, r_{xz} по формуле

$$R = \sqrt{\frac{r^2_{yx} + r^2_{yz} - 2r_{yx}\, r_{yz}\, r_{xz}}{1 - r^2_{xz}}}$$

Например, число пар (n) гибридов кукурузы составляет 30, а показателей (k) – 3. Коэффициенты корреляции для парных связей будут такими:

1) между урожайностью зерна (У) и массой початков молочно-восковой спелости (X) $r_{yx} = 0{,}79$;
2) между урожайностью зерна (У) и массой листьев (Z) $r_{yz} = 0{,}70$;
3) между массой початков (X) и массой листьев (Z) $r_{xz} = 0{,}82$.

Подставив эти коэффициенты в формулу, определяем коэффициент множественной корреляции

$$R = \sqrt{\frac{0{,}79^2 + 0{,}7^2 - (2 \cdot 0{,}79 \cdot 0{,}7 \cdot 0{,}82)}{1 - 0{,}82^2}} = 0{,}787.$$

Как видно, связь сильная (коэффициент корреляции изменяется в пределах 0,66-0,99).

Для оценки достоверности связи вычисляют критерий Фишера по формуле

$$F_R = \frac{R \cdot (n-k)}{(1-R^2)\cdot(k-1)} = \frac{0,787 \cdot (30-3)}{(1-0,787^2)\cdot(3-1)} = 27,9.$$

Теоретические значения критерия Фишера находят для числа степеней свободы $v_{\text{У}} = 2$ и $v_Z = n - 3 = 30 - 3 = 27$. Эти критерии составляют $F_{0,95} = 3,37$ и $F_{0,99} = 5,53$.

ВЫВОД: так как $F_R = 27,9$, что значительно превышает $F_{0,95} = 3,37$ и $F_{0,99} = 5,53$, то исследуемые связи являются высокодостоверными.

Для линейной множественной зависимости между тремя показателями **У, X** и **Z** уравнение регрессии имеет такой вид:

$$\mathbf{У = a + b_1X + b_2Z,}$$

где **У** – зависимая переменная,

a – общее начало отсчёта,

$\mathbf{b_1}$ и $\mathbf{b_2}$ **–** коэффициенты частной регрессии,

X и **Z –** независимые переменные.

Значения **a,** $\mathbf{b_1}$**,** $\mathbf{b_2}$ вычисляют методом наименьших квадратов.

7. ПРИМЕНЕНИЕ ЭВМ ДЛЯ ПРОВЕДЕНИЯ ИССЛЕДОВАНИЙ И ОБРАБОТКИ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Для статистической обработки экспериментальных данных используют пакет программ «Прикладная статистика на компьютере» SNEDECOR (О.Д. Сорокин, 2009).

8. КОНТРОЛЬНАЯ РАБОТА

Для более углублённого изучения студентами раздела 3. «Основы статистического анализа результатов исследований» дисциплины «Основы научных исследований в агрохимии и агропочвоведении» запланирована контрольная работа, которая включает выполнение индивидуальных заданий по темам:

1) вариационный ряд, его значение и применение в исследованиях по агрохимии и агропочвоведении;

2) корреляционный анализ данных, его значение и применение;

3) дисперсионный анализ экспериментальных данных, значение и применени.

СЛОВАРЬ ТЕРМИНОВ И ВЫРАЖЕНИЙ

Абсолютная ошибка – разница между истинным и фактически наблюдаемым значениями.

Абстрагирование – мысленное выделение основного в объекте исследований, его наиболее существенных связей (**отождествление** – для создания понятий о системах и классах, **изолирование** – для выделения основного среди второстепенного, что наиболее важно в абстракции).

Аддитивность – независимое действие факторов.

Анализ – метод исследований, с помощью которого исследуемый объект мысленно или физически разделяют на составные части для детального изучения.

Аналогия – метод, благодаря которому знания об известных уже объектах, предметах и явлениях переносятся на другие, похожие на них.

Антогонизм – когда одни факторы угнетают действие других;

Блок – часть повторения, компактная группа нескольких делянок, опыта.

Вариант – изучаемый вид растений, сорт, удобрение, агротехнические приёмы и т.д.

Вариационный ряд – такой ряд данных, в которых указаны возможные значения варьирующего признака в порядке возрастания или убывания и соответствующие им частоты.

Вегетационный метод – исследование растений в искусственных условиях, в вегетационных сосудах, в теплицах, оранжереях, вегетационных домиках, в фитотронах при строго контролируемых условиях внешней среды сроком от нескольких дней до нескольких месяцев.

Вегетационные опыты с почвенными культурами приближаются к производственным условиям. Используются для определения содержания в почве доступных (усвояемых) растениям питательных веществ и т. д.

Вегетационно-полевой метод – исследование растений непосредственно в поле в металлических цилиндрах, т.е. в сосудах без дна. Этот метод – промежуточный между вегетационным и полевым.

Вероятность – мера объективной возможности события, отношение числа благоприятных случаев, обозначается буквой **Р.**

Выборка – часть объектов совокупности (выборочная совокупность), подвергнутых непосредственному учёту, измерению.

Выключка – часть учётной делянки, исключённая из учёта из-за случайных повреждений или ошибок, допущенных при проведении опыта.

Вычислительные эксперименты – компьютерные расчёты математических моделей и выбор из их множества оптимальных.

Генеральная совокупность (выборка) – вся группа объектов, подлежащая изучению.

Географические опыты проводят в разных почвенно-климатических зонах по единой методике, разработанной научным координационным центром, который координирует исследования, принимает отчёты, обобщает результаты исследований и даёт рекомендации.

Главная книга опыта содержит основную информацию из полевого журнала, всю программу исследований (тема с научным обоснованием, методы исследований, рабочие гипотезы или несколько конкурирующих), схему опыта с выделением контрольных вариантов, размеры опытных делянок, ширину защиток, схему опыта в виде чертежа, результаты всех учётов и наблюдений с основными статистическими показателями.

Гипотеза – научное предположение, истинное значение которого является неопределённым.

Годовой отчёт о научно-исследовательской работе содержит только средние арифметические данные каждого варианта опыта, а в приложении – данные по повторностям с соответствующей статистической обработкой. Главный раздел отчёта – выводы и рекомендации производству. Для внедрения их в производство составляют специальные акты. После завершения темы научных исследований готовят итоговый отчёт за все годы работы. По результатам исследований пишут статьи, рефераты, диссертации.

Дактиль-метод используют для сокращения площади под стандартными вариантами, в котором стандарт размещают через 2 делянки и стандарт занимает 1/3 часть площади опыта.

Дедукция – метод исследований, который позволяет с помощью анализа общих положений и фактов делать частные одиночные выводы.

Делянка опытная – элементарная единица опыта, часть площади опыта, имеющая определённые размер и форму, предназначенная для размещения отдельного варианта.

Делянка учётная – часть площади опытной делянки, предназначенная для учёта урожая (без боковых и поперечных защиток).

Демонстрационные опыты – в них площадь опытных делянок обычно в 2 раза больше, чем в полевых опытах научных учреждений – 200-400 м^2, что необходимо для максимальной механизации производственных процессов.

Дисперсионный анализ – метод анализа результатов эксперимента, заключающийся в разложении общей изменчивости результативного признака, например, урожайности, на части – компоненты, соответствующие повторениям, вариантам, ошибкам случайного порядка и т.д. Значимость действия и взаимодействия изучаемых факторов оценивают по **F** и **НСР** $_{\mathbf{05}}$**.**

Дисперсия – статистический параметр (характеристика), представляющий средний квадрат отклонений отдельных значений совокупности от генеральной (выборочной) средней.

Длительные опыты ведут более 50 лет в отдельных институтах, почвенно-климатических зонах, краях, областях и др.

Дополнительная документация: лабораторный журнал; рабочая тетрадь; таблицы разных форм для всесторонних анализов; компьютерные графики, диаграммы и рисунки и т.п.

Достоверность опыта методическая – это чёткое соблюдение всех методических требований.

Достоверность опыта статистическая состоит в определении достоверности (существенности) разницы между средними арифметическими значениями **(x)**, корреляции **(r)**, регрессии **(Rxy)** и др. с помощью статистических критериев **(t, F)** и наименьших существенных разностей **НСР**.

Единичные опыты проводят в разных географических пунктах по схеме, созданной отдельными исследователями или их группами, они позволяют обобщать результаты в пределах района, области, края и в отдельных почвенно-климатических зонах.

Защитная полоса, защитка – краевые (боковые и поперечные) части делянок, которые не подвергаются учёту и служат для исключения влияния соседних вариантов, для предохранения учётной части делянок от случайных повреждений, для разворота машин, орудий и т.п.

Значимость (существенность) – мера объективной возможности (риск) сделать ошибочное заключение при оценке результатов опыта. При оценке результатов полевого опыта принято опираться на 5%-ный уровень значимости, при котором риск сделать ошибочное заключение составляет 5%. При более строгой оценке принимают 1%-ный уровень значимости.

Изменчивость – вариабельность, вариация индивидуальных значений признаков X около среднего значения $\bar{x}$. Основной мерой изменчивости являются дисперсия S^2 и стандартное отклонение S.

Контроль (стандарт) – один или несколько вариантов, с которыми сравнивают опытные варианты.

Корреляционный анализ – статистический метод определения тесноты и формы связи между признаками.

Корреляция – взаимосвязь между признаками, заключающаяся в том, что средняя величина значений одного признака меняется в зависимости от изменения другого признака.

Коэффициент вариации (изменчивости) – относительный показатель изменчивости признака, представляющий отношение стандартного отклонения S к средней арифметической, выраженное в процентах. Обозначается буквой V.

Коэффициент корреляции – статистический показатель тесноты (силы) связи. Обозначается буквой r.

Коэффициент регрессии – R_{yx} – число, показывающее, в каком направлении и на какую величину изменяется в среднем зависимая переменная У (результативный признак) при изменении независимой переменной Х на единицу измерения.

Инверсия – метод необычного изучения объектов, явлений (под определённым углом и даже с противоположной стороны); соединение несовместимого, деление неделимого – отказ от общепринятых взглядов и приёмов.

Индукция – метод исследований, с помощью которого суждения ведут от фактов к конкретным выводам.

История полей – в ней указывают, где и какие культуры выращивались на конкретном поле в предшествующие годы, после каких предшественников; историю поля желательно знать за 2-3 года до закладки опыта, а ещё лучше – за всю ротацию севооборота.

Качественная изменчивость – варьирование, когда различия между вариантами выражаются качественными показателями, которые одни варианты имеют, а другие нет (цвет, вкус, форма изучаемого объекта).

Качественные эксперименты – в них учитывают наличие или отсутствие того, или иного качественного показателя (повреждённые или неповреждённые вредителями, или болезнями растения).

Количественная изменчивость – такая, в которой различия между вариантами выражаются количеством (масса урожая, процент сахара, кислот, витаминов, крахмала или белка в урожае; размеры растений, содержание питательных элементов в почве), т.е. всё, что имеет массу, размер, объём и т.п.

Количественные эксперименты – в них учитывают количественные показатели (рост растений, их урожайность, содержание белка в зерне пшеницы, сахара в корнях сахарной свёклы).

Конкретизация – метод исследований, с помощью которого от абстрактного переходят к конкретному.

Контрольный вариант – вариант, с которым сравнивают все остальные варианты.

Корреляция – взаимное соотношение показателей в опыте, их зависимость между собой, когда изменение одного признака вызывает изменения другого; чем больше эти изменения, тем сильнее корреляционная связь.

Корреляция прямолинейная – когда с увеличением одного из показателей второй также увеличивается.

Корреляция криволинейная – когда с увеличением одного признака (например, доз удобрений) значение второго сначала увеличивается (например, урожайность), затем стабилизируется на одном уровне, а потом снижается.

Корреляционный анализ – статистический метод определения тесноты и формы связи между признаками.

Корреляционный коэффициент (коэффициент корреляции) – показатель тесноты линейной связи двух признаков; корреляция сильная при **$r > 0,7$** и слабая при **$r < 0,5$.**

Коэффициент вариации (изменчивости) – относительный показатель изменчивости признака, представляющий отношение стандартного отклонения S к средней арифметической, выраженное в процентах. Обозначается буквой V.

Коэффициент корреляции – статистический показатель тесноты (силы) связи. Обозначается буквой r.

Коэффициент регрессии – R_{yx} – число, показывающее, в каком направлении и на какую величину изменяется в среднем зависимая переменная У (результативный признак) при изменении независимой переменной X на единицу измерения.

Краткосрочные опыты проводят в течение 3-10 лет, обычно на протяжении ротации севооборота; краткосрочными являются опыты, которые ведут студенты для написания дипломных (выпусных) работ или аспиранты во время подготовки диссертации.

Криволинейная корреляция – значения признаков **X** и **У** изменяются сначала в одном направлении, а затем в противоположных (например, постоянное увеличение доз фактора **X** (азотные или другие удобрения, влажность почвы, её плотность и т.п.) урожайность **У** сначала возрастает, затем стабилизируется, а после дальнейшего увеличения признак X снижается.

Лабораторный метод используют для анализа растений и среды их обитания в лабораторных условиях для изучения взаимодействия растений с внешней средой, обмена веществ в растениях, оценки качества урожая, исследования физических, химических, микробиологических свойств почвы и т.д.

Лабораторный опыт проводят в лабораторных условиях без опытных растений.

Латинский квадрат – схема рендомизированного (случайного) размещения вариантов в опыте, в котором делянки располагают рядами и столбцами (4х4, 5х5, 6х6 и т.д.). В каждом ряду и столбце должен быть полный набор вариантов схемы (повторения), и, следовательно, в латинском квадрате число повторений равно числу вариантов, а общее число делянок равно квадрату числа вариантов.

Латинский прямоугольник – схема рендомизированного (случайного) размещения вариантов в опыте. В основе лежит латинский квадрат, который и определяет повторность опыта, число рядов и столбцов. Число вариантов должно быть кратным повторности (4х3х3; повторность n=4, число вариантов l=12).

Лизиметрический метод – исследование растений и свойств почвы в поле для изучения баланса влаги и элементов питания, проводят в очень больших сосудах – лизиметрах, которые периодически взвешивают.

Линейная корреляция – когда с увеличением признака **Х** соответственно увеличивается второй признак **У** (например, площадь листьев возрастает с увеличением их длины).

Мало управляемые факторы – температура воздуха и почвы, свет и т.п.

Математическая статистика – это один из разделов математики, позволяет делать умозаключения о всей (генеральной) совокупности на основе наблюдений над выборочной совокупностью, или выборкой. Все статистические методы основаны на **теории вероятностей**.

Мелкоделяночный полевой опыт – очень близок к полевому опыту, но отличается от него тем, что площади опытных делянок имеют размер в несколько квадратных метров: 4, 8, 10; а все работы на таких делянках выполняют вручную, количество растений незначительно, а их агротехника – не типичная для производственных условий.

Метод – упорядоченная деятельность исследователя, направленная на получение новых знаний. В агрономии используют **общенаучные** и **конкретно-научные (специальные)** методы.

Метод расщеплённых (сложных) делянок – эксперимент, в котором делянки одного опыта используют как блоки для другого. Делянки первого порядка расщепляют на делянки второго порядка, а последние – на более мелкие делянки третьего порядка. Метод расщеплённых делянок с рендомизированным размещением вариантов используют для закладки многофакторных опытов.

Метод рендомизированных (случайных) повторений – эксперимент, в котором варианты по делянкам размещены в случайном порядке по таблице случайных чисел или по жребию. Это наиболее распространённый метод размещения вариантов.

Методика полевого опыта – включает следующие элементы: число вариантов, площадь делянок, их форма и направление, повторность, система размещения вариантов, повторений и делянок на территории, метод учёта урожая, организация опыта во времени, а также метод статистического анализа данных.

Методы размещения вариантов в опытах – **рендомизированные (случайные)**, т.е. выбранные по жребию; - **систематические –** варианты размещаются в последовательности, которая указана в схеме опыта; - **стандартные,** когда контрольный вариант размещается возле опытного.

Метод рендомизированных расщеплённых делянок – это размещение вариантов фактора первого порядка на основных делянках, а факторов второго и последующих порядков – на субделянках, на которые расщепляют основные делянки.

Метод смешивания – наиболее эффективен в многофакторных опытах, когда все варианты делят на несколько равноценных групп так, чтобы разницу между этими группами составляли взаимодействия высшего порядка, которые меньше интересуют исследователя, чем факторы низшего порядка. Но при этом теряется информация о взаимодействии факторов высших порядков (недостаток метода смешивания).

Многолетние опыты проводят в течение 11-50 лет в научно-исследовательских учреждениях или высших учебных заведениях на специально выделенных участках (стационарах).

М**ножественная корреляция** – когда результат опыта зависит одновременно от нескольких показателей или факторов.

Многофакторный опыт – включает одновременно несколько факторов (различные площади питания, сроки посева, несколько сортов и т.п.); это более сложные опыты, но они дают больше информации и поэтому имеют большую научную и практическую ценность.

Моделирование – метод исследования объектов, процессов и явлений на их моделях.

Наблюдение – это целенаправленное сосредоточение внимания исследователя на явлениях, происходящих в эксперименте или на явлениях природы, их количественная и качественная регистрация.

Наименьшая существенная разница (НСР) – величина, указывающая границу возможных случайных отклонений в эксперименте; это минимальная разность в урожаях между средними, которая в данном опыте признаётся существенной при 5%-ном ($НСР_{05}$) или 1%-ном ($НСР_{01}$) уровне значимости.

Наука – это определённый предмет изучения или дисциплина, представляющая совокупность систематизированных знаний относительно своих объектов и терминологии.

Научное исследование – изучение конкретного объекта, явления или предмета для раскрытия закономерностей его возникновения и развития.

Научный отдел – основная структурная часть опытной станции или института; в его состав входят научные лаборатории, которые занимаются исследованиями по конкретной тематике.

Научно-исследовательские институты – это учреждения, которые разрабатывают теоретические проблемы сельскохозяйственной науки и практические рекомендации для развития определённых отраслей агрономии. Институты могут быть как зональными, так и отраслевыми.

Непрерывная изменчивость – когда, значения вариантов выражаются мерами объёма, длины, массы и т.д., между которыми мыслимы любые переходы с неограниченным числом возможных значений; всё зависит от степени точности, принимаемой для характеристики данного количественного признака.

Неуправляемые факторы – атмосферные осадки, зимние морозы – нарушают процесс воспроизводимости опыта.

НСР – наименьшая существенная разность, предел случайных отклонений или критическая разность (если прибавка от изучаемого варианта больше НСР, значит этот вариант достоверен).

Область определения фактора – совокупность всех значений, которые может принимать данный фактор; они могут быть количественными (дозы удобрений, глубина вспашки, площадь питания растений) и качественными (формы удобрений, сорта, разные пестициды).

Обобщение – метод, с помощью которого мысленно переходят от отдельных факторов, явлений и процессов к отождествлению в мыслях; от одного понятия, суждения к более общему. Так обобщают результаты исследований для каждого повторения, затем для всего опыта, конкретного хозяйства, группы хозяйств, которые находятся в аналогичных почвенно-климатических условиях**.**

Объём выборки – число элементов в генеральной совокупности и выборке.

Однофакторный опыт – в нём изучают только один фактор (только различные площади питания, только сроки посева или же несколько сортов растений, но на одном агротехническом фоне).

Опорный пункт – это научное подразделение опытной станции или института, которое создаётся на производстве.

Опыт, эксперимент в агрономии – это создание различных условий для исследуемых растений с целью выявления наиболее эффективных вариантов в процессе учётов и наблюдений.

Опытная делянка – это земельная площадь прямоугольной формы определённого размера, на которой изучают только один из вариантов опыта – агроприём, технологию, сорт и т.д.

Опытное дело в агрономии – научно-исследовательская работа, разработка теории и практики повышения продуктивности сельскохозяйственных культур, качества продукции при минимальных затратах труда и средств.

Опытные поля призваны проводить многолетние стационарные полевые опыты для выявления лучших приёмов возделывания сельскохозяйственных культур в конкретных почвенно-климатических условиях (эффективность минеральных и органических удобрений, типы севооборотов, способы борьбы с эрозией почв, технологии выращивания сельскохозяйственных культур).

Опыты-пробы – закладывают на производственных посевах, где выделяют полосы шириной в один проход жатки или комбайна. Длина таких делянок должна быть в 5-10 раз больше ширины.

Опыты по учёту эффективности новых агроприёмов в производстве – выделяют в виде контрольных полос, ширина которых должна соответствовать ширине прохода уборочного агрегата, а длина – длине загонок. Общая площадь каждой из этих полос составляет до 3 га.

Опытные станции осуществляют научную разработку агротехнических мероприятий в конкретных естественно-экономических условиях и дают рекомендации производству. Они ведут пропаганду достижений науки и передовой практики.

Основная документация при проведении полевого опыта – полевой журнал (дневник научного работника), главная книга опыта, рабочая программа и отчёт о научно-исследовательской работе.

Относительная ошибка опыта (наблюдения) – это ошибка опыта, выраженная в % по отношению к среднему арифметическому значению и обозначается **Sx%**.

Ошибка опыта (наблюдения) – разница между действительным значением исследуемого показателя и результатами исследований. Эту ошибку опыта (наблюдения) выражают в тех же самых единицах, что и изучаемый показатель, обозначают **Sx.**

Парный метод – когда опытные делянки делят поперёк на маленькие деляночки (парцеллы).

Повторность опыта – число делянок в каждом опыте с одинаковым содержанием вариантов, т.е. с одинаковыми агротехническими приёмами или сортами растений и т.п..

Полная рендомизация вариантов – рендомизированное размещение всех вариантов опыта без предварительного выделения повторений; её используют, когда индивидуальное варьирование растений превышает варьирование плодородия почвы (это характерно для многолетних растений, особенно древесных); когда число вариантов и повторностей в опыте невелико (3-4); когда размеры опытных делянок и площадь под опытом небольшие и варьирование плодородия почвы незначительное.

Полевой метод – проведение полевого опыта (эксперимента) – основной метод в научной агрономии, он позволяет связать теоретические исследования с практическими; проводят в полевой обстановке на специально выделенном участке в целях установления влияния факторов жизни, условий или приёмов возделывания на урожайность сельскохозяйственных растений и его качество.

Прерывистая (дискретная) изменчивость – различия между вариантами, которые выражаются целыми числами, между которыми нет и не может быть переходов (например, число растений на квадратном метре, число зёрен в колосе и т.д.).

Прикладные исследования в агрономии направлены на изучение факторов жизни растений и взаимосвязей между растениями и средой, на создание перспективных сортов и гибридов.

Принцип единственного логического развития – можно изменять лишь изучаемый фактор при строгом постоянстве всех остальных условий опыта (например, в однофакторном опыте изучают продуктивность яровой пшеницы в зависимости от нормы высева: 4, 5 и 6 млн. зёрен на 1 га; согласно принципу единственного различия в этом опыте изменяют лишь норму высева, остальные элементы агротехники – предшественник, удобрения, обработка почвы, сроки, глубина и способ посева, используемые агрегаты, уход за посевами, метод уборки, должны быть одинаковыми).

Производственные опыты проводят на всей площади севооборота, на площади полевой бригады и даже целого хозяйства или административного района.

Рабочая программа – её составляют на весь период исследований, т.е. на несколько лет; а на каждый год исследований также составляют планы научно-исследовательской работы, один из них – календарный план последовательного проведения всех работ по срокам на протяжении всего года.

Разведывательные (временные) опыты – проводят на протяжении 1-2 лет для выявления тех агроприёмов или сортов растений, которые необходимо изучать в последующих опытах (например, рекогносцировочные посевы для выявления степени изменения плодородия почвы на месте будущего опыта).

Ранжирование – такое упорядочение вариационного ряда, т.е. расположение вариант в порядке возрастания (или убывания) признака.

Регрессия – степень и характер изменения одного из показателей в опыте на единицу измерения другого (например, увеличение или уменьшение массы урожая на 100 кг внесённых удобрений); регрессия обозначается **Rxy.**

Рекогносцировочный или **разведовательный посев** – сплошной посев перед закладкой опыта с учётом урожая для выявления вида и степени варьирования почвенного плодородия на всей площади будущего опыта.

Рендомизация – размещение вариантов по экспериментальным единицам, основанное на жребии или случайности.

Рендомизированный блок – план опыта со случайным размещением вариантов по всем повторениям.

Рендомизированный латинский прямоугольник – случайное размещение всех вариантов в пределах каждого ряда и каждого отдельного блока; метод эффективен тогда, когда плодородие почвы изменяется не только в двух взаимно перпендикулярных направлениях, но и по диагонали.

Рендомизированные повторения – случайное размещение всех вариантов опыта в пределах отдельных повторений; используется, если в пределах повторения (блока) варьирование плодородия почвы минимальное, а между повторениями (блоками) оно колеблется в большей мере, если различия между блоками отсутствуют, данный метод малоэффективен.

Репрезентативность – соответствие выборки природе генеральной совокупности и наличие достаточного объёма.

Синергизм – когда одни факторы усиливают действие других.

Синтез – соединение разделённых и проанализированных частей исследуемого объекта или нескольких объектов в единое целое.

Системный анализ – тесное переплетение элементов науки и практики (например, в экспериментах можно исключать влияние побочных факторов, выделяя исследуемое явление, можно вводить новые факторы, результаты исследований можно многократно воспроизводить).

Систематический (последовательный) метод – кода варианты размещают в последовательности, указанной в схеме опыта, или же по другой системе, но во всех повторениях одинаково; он эффективен, если нет закономерного (систематического) изменения плодородия почвы.

Случайный метод (или рендомизация) – когда место вариантов определяют по таблице случайных чисел или по жребию.

Стандарт – лучший сорт среди районированных и наиболее распространённых, с которым сравнивают остальные изучаемые сорта.

Стандартный метод – рядом с каждым опытным вариантом размещают контроль (стандарт), эффективен тогда, когда плодородие почвы значительно варьирует (например, дерново-подзолистые, солонцовые и другие почвы).

Суждение – такая форма мышления, когда утверждают, либо отрицают существование явления, процесса, оно может быть **объективным** или **ошибочным**.

Сущность системного подхода – исследование объекта как системы (например, исследование живого организма с его составными частями, внутренними и внешними связями и т.д.).

Схема опыта – это перечень логично подобранных вариантов с определёнными контролями (стандартами), объединённых конкретной темой.

Схематический план опыта – представление основных параметров опыта: варианты, повторность, размеры (площадь) делянок и защитных полос, размещение и привязка их на местности.

Теория – это система обобщённых знаний, объяснение определённых явлений действительности.

Теория вероятностей – наука, изучающая общие закономерности в массовых случайных явлениях различной природы; применяется везде, где приходится иметь дело с планированием экспериментов и обследований, с оценкой параметров и проверкой гипотез, с принятием решений при изучении сложных систем.

Типичность – одно из основных условий опыта, его нарушение обесценивает опыт и приводит к тому, что его результаты не могут быть рекомендованы производству (поэтому опыты необходимо проводить в таких условиях, которые соответствовали бы природной зоне, почвам, особенностям выращиваемой культуры и сорту, уровню механизации, глубине залегания грунтовых вод, организационно-экономическим условиям и т.п.).

Точность опыта – условное понятие, которое определяется величиной относительной ошибки опытных средних (V): до 3% - высокая, до 5% - допустимая, более 7% - неприемлемая точность.

Точные сравнительные опыты – в них ширина делянки должна быть кратна ширине прохода почвообрабатывающих, посевных и уборочных агрегатов, чтобы полнее механизировать наиболее трудоёмкие процессы: в посевах с культурами сплошного способа посева (однолетние и многолетние травы, зерновые) она составляет 8-16 м; с пропашными культурами (кукуруза, картофель) – 5-10 м, а их общая площадь 500-2000 м^2.

Умозаключение – такая форма мышления, когда из одного или нескольких связанных между собой суждений выводят новые знания.

Управляемые факторы – сорт, удобрение, обработка почвы, схема посева и т.п.

Уравнительный посев – сплошной посев культуры с целью оценки почвенного покрова и его выравнивания за 2-4 года до закладки опыта.

Учётная часть делянки находится внутри неё, где проводят все учёты и наблюдения, она является элементарной учётной единицей в опытах.

Фактор – переменная величина, принимающая в некоторый момент определённое значение, это способ воздействия на объект исследования.

Формализация – метод изучения объектов с помощью отдельных элементов их форм, которые отображают содержание объекта исследования (его чаще всего применяют с использованием математики, приводя доказательства в виде последовательных формул).

Фундаментальные исследования направлены на открытие и изучение новых явлений и законов природы.

Шаг эксперимента – разница между последующей и предыдущей дозами фактора (дозы удобрений, нормы высева и др.).

Ширина продольных защитных полос (защиток) составляет 1-1,5 м. Более широкими делают защитки вокруг всего опыта (для защиты от наезда транспорта, дорожной пыли, потрав животных).

Экспериментальная единица – место (делянка, сосуд, пробирка), одно растение для наложения одного варианта.

Эксперимент или **опыт** – метод научного исследования, осуществляемый на основе тщательно разработанного плана с целью получения новых знаний по чётко сформулированной проблеме.

Экспедиционный метод – используют для изучения и обобщения агрономических вопросов непосредственно на производстве с помощью обследования посевов культур (и сортов).

Ямб-метод – когда стандарт размещается через одну делянку

СПИСОК ЛИТЕРАТУРЫ

Основная литература

1. Овчаров А. О. Методология научного исследования: Учебник / А.О. Овчаров, Т.Н. Овчарова. – 2-е изд., испр. и доп. – Москва: ИНФРА-М, 2022. – 310 с. Доп. материалы [Электронный ресурс]. – (Высшее образование: Магистратура). – DOI 10.12737/1846123. – ISBN 978-5-16-017366-5. – Текст: электронный. – URL: https://znanium.com/catalog/product/1846123.

Дополнительной литература

1. Кирюшин Б.Д. Основы научных исследований в агрономии: учебник. – Санкт-Петербург: Квадро, 2013. – 408 с.

2. Дмитриев Е.А. Математическая статистика в почвоведении. – М.: Изд-во МГУ, 1999.

3. Моисейченко В.Ф. Основы научных исследований в агрономии / В.Ф. Моисейченко, М.Ф. Трифонова, А.Х. Заверюха, В.Е. Ещенко. – М.: Колос, 1996. – 336 с.

4. Доспехов Б.А. Методика полевого опыта (с основами статистической обработки результатов исследований) – М.: Агропромиздат, 1985. – 351 с.

5. Журбицкий З.И. Теория и практика вегетационного метода. – М.: Наука, 1968. – 265 с.

6. Кирюшин В.И. Агрономическое почвоведение. – М.: КолосС, 2010. – 687 с.

7. Овчаров А. О.Методология научного исследования: Учебник / А.О. Овчаров, Т.Н. Овчарова. - М.: НИЦ ИНФРА-М, 2014. - 304 с. (ЭБС)

8. Основы научных исследований в агрономии: Рабочая тетрадь для лабораторно-практических занятий // Р.Р. Галеев / Новосиб. гос. аграр. ун-т,

Новосибирск: Агро-Сибирь, 2011. – 38 с.

9. Перегудов В.Н. Планирование многофакторных опытов с удобрениями и математическая обработка результатов. – М.: Колос, 1978. – 206 с.

10. Пискунов А.С. Методы агрохимических исследований. – М.: КолосС, 2004. – 312 с.

11. Сорокин О.Д. Прикладная статистика на компьютере. – Новосибирск: ГУП РПО СО РАСХН, 2009. – 222 с.

13. Средства обеспечения дисциплины: программные компьютерные комплексы: STATISTICA, ППП STRAZ ПК, STATGRAFICS plus for windows и др.

14. Сайты Интернета: char.ru, dony.ru, club-book.net, bookland.ru, goraknig.org, bibliolink.ru и др.

ПРИЛОЖЕНИЯ

1. ЗНАЧЕНИЯ КРИТЕРИЯ *t* НА 5, 1 И 0,1%-НОМ УРОВНЯХ ЗНАЧИМОСТИ

Число степеней свободы	Уровень значимости			Число степеней свободы	Уровень значимости		
	0,05	0,01	0,001		0,05	0,01	0,001
1	12,71	63,66	—	18	2,10	2,88	3,92
2	4,30	9,93	31,60	19	2,09	2,86	3,88
3	3,18	5,84	12,94	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	6,86	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	3,06	4,32	29	2,05	2,76	3,66
13	2,16	3,01	4,22	30	2,04	2,75	3,65
14	2,15	2,98	4,14	50	2,01	2,68	3,50
15	2,13	2,95	4,07	100	1,98	2,63	3,39
16	2,12	2,92	4,02				
17	2,11	2,90	3,97				

2. ЗНАЧЕНИЯ КРИТЕРИЯ F НА 5%-НОМ УРОВНЕ ЗНАЧИМОСТИ (ВЕРОЯТНОСТЬ 95 %)

Степень свободы для меньшей дисперсии (знаменателя)	Степень свободы для большей дисперсии (числителя)													
	1	2	3	4	5	6	7	8	9	10	12	24	50	100
1	161	200	216	225	230	234	237	239	241	242	244	249	452	253
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,41	19,45	19,47	19,49
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,74	8,64	8,58	8,56
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,77	5,70	5,66
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,68	4,53	4,44	4,40
6	5,99	5,14	4,76	4,53	4,39	4,27	4,21	4,15	4,10	4,06	4,00	3,84	3,75	3,71
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,57	3,41	3,32	3,28
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,28	3,12	3,03	2,98
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,07	2,90	2,80	2,76
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,91	2,74	2,64	2,59
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,79	2,61	2,50	2,45
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,69	2,50	2,40	2,35
13	4,64	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,60	2,42	2,32	2,26
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,53	2,35	2,24	2,19
15	4,54	3,67	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,48	2,29	2,18	2,12
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,24	2,13	2,07
17	4,45	3,59	3,20	2,96	2,81	2,69	2,62	2,55	2,50	2,45	2,38	2,19	2,08	2,02
18	4,41	3,55	3,18	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,15	2,04	1,98
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,31	2,11	2,00	1,94
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,28	2,08	1,96	1,90
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,05	1,93	1,87
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,23	2,03	1,91	1,84
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,20	2,00	1,88	1,82
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,18	1,98	1,86	1,80
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,25	2,24	2,16	1,96	1,84	1,77
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,25	2,22	2,15	1,95	1,82	1,76
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,12	1,91	1,78	1,72
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,12	2,09	1,89	1,76	1,69
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,00	1,79	1,66	1,59
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,95	1,74	1,60	1,52
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,85	1,63	1,48	1,39

3. ЗНАЧЕНИЕ КРИТЕРИЯ F НА 1%-НОМ УРОВНЕ ЗНАЧИМОСТИ (ВЕРОЯТНОСТЬ 99 %)

Степень свободы для меньшей дисперсии (знаменателя)	Степень свободы для большей дисперсии (числителя)													
	1	2	3	4	5	6	7	8	9	10	12	24	50	100
1	4042	4999	5403	5625	5764	5889	5928	5981	6022	6056	6106	6234	6302	6334
2	98,49	99,01	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40	99,42	99,46	99,48	99,49
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,05	26,60	26,35	26,23
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,37	13,93	13,69	13,57
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,89	9,47	9,24	9,13
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,31	7,09	6,99
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,47	6,07	5,85	5,75
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,67	5,28	5,06	4,96
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,11	4,73	4,51	4,41
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,71	4,33	4,12	4,01
11	9,85	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,40	4,02	3,80	3,70
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,16	3,78	3,56	3,46
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,59	3,37	3,27
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,43	3,21	3,11
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,29	3,07	2,97
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,78	3,69	3,61	3,54	3,18	2,96	2,86
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,45	3,08	2,86	2,76
18	8,28	6,01	5,09	4,58	4,25	4,01	3,85	3,71	3,60	3,51	3,37	3,00	2,78	2,68
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,68	3,52	3,43	3,30	2,92	2,70	2,63
20	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37	3,23	2,86	2,63	2,53
21	8,02	5,78	4,87	4,37	4,04	3,81	3,65	3,51	3,40	3,31	3,17	2,80	2,58	2,47
22	7,94	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,75	2,53	2,42
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,70	2,48	2,37
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,25	3,17	3,03	2,66	2,44	2,33
25	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,21	3,13	2,99	2,62	2,40	2,29
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,17	3,09	2,96	2,58	2,36	2,25
28	7,64	5,45	4,57	4,07	3,76	3,53	3,36	3,23	3,11	3,03	2,90	2,52	2,30	2,18
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,06	2,98	2,84	2,47	2,24	2,13
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,88	2,80	2,66	2,29	2,05	1,94
50	7,17	5,06	4,20	3,72	3,41	3,18	3,02	2,88	2,78	2,70	2,56	2,18	1,94	1,81
100	6,90	4,82	3,98	3,51	3,20	2,99	2,82	2,69	2,59	2,51	2,36	1,98	1,73	1,59

4. Значения критерия τ для 5%-ного и 1%-ного уровней значимости

n	τ		n	τ	
	0,01	0,05		0,01	0,05
4	0,991	0,995	13	0,502	0,395
5	0,916	0,807	16	0,472	0,369
6	0,805	0,689	18	0,449	0,349
7	0,740	0,610	20	0,430	0,334
8	0,683	0,554	22	0,414	0,320
9	0,635	0,512	24	0,400	0,309
10	0,597	0,477	26	0,389	0,299
11	0,566	0,460	28	0,378	0,291
12	0,541	0,428	30	0,369	0,283

5. УГЛЫ, СООТВЕТСТВУЮЩИЕ ПРОЦЕНТАМ: УГОЛ-АРКСИНУС √ПРОЦЕНТ

%	Десятые доли процента									
	0	1	2	3	4	5	6	7	8	9
0	0,0	1,8	2,6	3,1	3,6	4,0	4,4	4,8	5,1	5,4
1	5,7	6,0	6,3	6,6	6,8	7,0	7,3	7,5	7,7	7,9
2	8,1	8,3	8,5	8,7	8,9	9,1	9,3	9,5	9,6	9,8
3	10,0	10,1	10,3	10,5	10,6	10,8	10,9	11,1	11,2	11,4
4	11,5	11,7	11,8	12,0	12,1	12,2	12,4	12,5	12,7	12,8
5	12,9	13,0	13,2	13,3	13,4	13,6	13,7	13,8	13,9	14,1
6	14,2	14,3	14,4	14,5	14,6	14,8	14,9	15,0	15,1	15,2
7	15,3	15,4	15,6	15,7	15,8	15,9	16,0	16,1	16,2	16,3
8	16,4	16,5	16,6	16,7	16,8	17,0	17,1	17,2	17,3	17,4
9	17,5	17,6	17,7	17,8	17,8	18,0	18,0	18,2	18,2	18,3
10	18,4	18,5	18,6	18,7	18,8	18,9	19,0	19,1	19,2	19,3
11	19,4	19,5	19,6	19,6	19,7	19,8	19,9	20,0	20,1	20,2
12	20,3	20,4	20,4	20,5	20,6	20,7	20,8	20,9	21,0	21,0
13	21,1	21,2	21,3	21,4	21,5	21,6	21,6	21,7	21,8	22,0

Продолжение

%	Десятые доли процента									
	0	1	2	3	4	5	6	7	8	9
14	22,1	22,1	22,1	22,2	22,3	22,4	22,5	22,6	22,6	22,7
15	22,8	22,9	23,0	23,0	23,1	23,2	23,3	23,3	23,4	23,5
16	23,6	23,7	23,7	23,8	23,9	24,0	24,0	24,1	24,2	24,3
17	24,4	24,4	24,5	24,6	24,6	24,7	24,8	24,9	25,0	25,0
18	25,1	25,2	25,2	25,3	25,4	25,5	25,6	25,6	25,7	25,8
19	25,8	25,9	26,0	26,1	26,1	26,2	26,3	26,4	26,4	26,5
20	26,6	26,6	26,7	26,8	26,9	26,9	27,0	27,1	27,1	27,2
21	27,3	27,4	27,4	27,5	27,6	27,6	27,7	27,8	27,8	27,9
22	28,0	28,0	28,1	28,2	28,2	28,3	28,4	28,4	28,5	28,6
23	28,7	28,7	28,8	28,9	28,9	29,0	29,1	29,1	29,2	29,3
24	29,3	29,4	29,5	29,5	29,6	29,7	29,7	29,8	29,9	29,9
25	30,0	30,1	30,1	30,2	30,3	30,3	30,4	30,5	30,5	30,6
26	30,7	30,7	30,8	30,9	30,9	31,0	31,0	31,0	31,2	31,2
27	31,2	31,3	31,4	31,5	31,6	31,6	31,7	31,8	31,8	31,9
28	32,0	32,0	32,1	32,1	32,2	32,3	32,3	32,4	32,5	32,5
29	32,6	32,6	32,7	32,8	32,8	33,9	33,0	33,0	33,1	33,2
30	33,2	33,3	33,3	33,4	33,5	33,5	33,6	33,6	33,7	33,8
31	33,8	33,9	34,0	34,0	34,1	34,1	34,2	34,3	34,3	34,3
32	34,4	34,5	34,6	34,6	34,7	34,8	34,8	34,9	35,0	35,0
33	35,1	35,1	35,2	35,2	35,3	35,4	35,4	35,5	35,6	35,6
34	35,7	35,7	35,8	35,9	35,9	36,0	36,0	36,1	36,2	36,2
35	36,1	36,3	36,4	36,5	36,5	36,6	36,6	36,7	36,8	36,8
36	36,9	36,9	37,0	37,0	37,1	37,2	37,2	37,3	37,4	37,4
37	37,5	37,5	37,6	37,6	37,7	37,8	37,8	37,9	37,9	38,0
38	38,1	38,1	38,2	38,2	38,3	38,4	38,4	38,5	38,5	38,6
39	38,6	38,7	38,8	38,8	38,9	38,9	39,0	39,1	39,1	39,2
40	39,2	39,3	39,4	39,4	39,5	39,5	39,6	39,6	39,7	39,8
41	39,8	39,9	39,9	40,0	40,0	40,1	40,2	40,2	40,3	40,3
42	40,4	40,5	40,5	40,6	40,6	40,7	40,7	40,8	40,9	40,9
43	41,0	41,0	41,1	41,2	41,2	41,3	41,3	41,4	41,4	41,5

ОГЛАЛЕНИЕ