

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ –
МСХА имени К.А. ТИМИРЯЗЕВА

Р.Р. Усманов

**ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ
В БИОТЕХНОЛОГИИ
(с расчетами в программе Excel)**

Учебно-практическое пособие

МОСКВА

РГАУ-МСХА имени К.А. Тимирязева

2022

УДК 58.084/.085:57.087.1(075.8)

ББК 28с

У 74

Основы научных исследований в биотехнологии (с расчетами в программе Excel)

У 74 **Усманов, Р. Р.** Основы научных исследований в биотехнологии (с расчетами в программе Excel): учебно-практическое пособие / Р.Р. Усманов; Российский государственный аграрный университет – МСХА имени К. А. Тимирязева. – Москва: РГАУ-МСХА имени К. А. Тимирязева, 2022. – 122 с. – Текст: электронный.

DOI: 10.34677/

В пособии изложен материал по выполнению практических и семинарских занятий по курсу «Основы научных исследований в биотехнологии» с методическими указаниями по применению программы Excel для статистической обработки результатов экспериментов в биотехнологии. Пособие составлено в соответствии с рабочей программой дисциплины и предназначено для студентов по направлению подготовки «Биотехнология».

Рекомендовано к изданию учебно-методической комиссией Института агробиотехнологии РГАУ-МСХА имени К.А. Тимирязева, протокол № 4 от 10 января 2022 года).

Usmanov, R. R. Fundamentals of scientific research in biotechnology (with calculations in Excel): an educational and practical manual / R.R. Usmanov; «Russian State Agrarian University – Moscow Timiryazev Agricultural Academy», 2022. – 122 s. – Text: electronic.

DOI: 10.34677/

The manual contains material on the implementation of practical and seminar classes on the course «Fundamentals of scientific research in biotechnology» with methodological guidelines for the use of Excel software for statistical processing of experimental results in biotechnology. The manual is compiled in accordance with the work program of the discipline and is intended for students in the field of training «Biotechnology».

© Усманов Р.Р., 2022

© ФГБОУ ВО РГАУ – МСХА
имени К.А. Тимирязева, 2022

ОГЛАВЛЕНИЕ

Введение	5
Работа 1. Выборочный метод в биотехнологических исследованиях	6
Основные статистические показатели данных наблюдений	6
Статистическая обработка данных биотехнологических исследований в программе «Excel»	10
Расчеты с использованием МАСТЕРА ФУНКЦИЙ	10
Расчеты с использованием надстройки Анализ данных	15
Работа 2. Группировка и графическое представление данных наблюдений	21
Группировка данных, расчет статических показателей, построение гистограммы в программе Excel	26
Работа 3. Оценка двух вариантов при количественной изменчивости признаков	31
А. Независимые выборки	33
Оценка двух вариантов при количественной изменчивости признаков для независимых выборок в программе Excel	35
В. Зависимые (сопряженные) выборки	45
Оценка средней разности при количественной изменчивости признаков для зависимых выборок в программе Excel	47
Работа 4. Оценка двух вариантов при качественной изменчивости признаков	49
Работа 5. Оценка соответствия между фактическими и теоретическими наблюдениями	52
Оценка соответствия между фактическими и теоретическими наблюдениями в программе Excel	54
Работа 6. Обработка данных с применением непараметрических критериев	60
Работа 7. Корреляционно-регрессионный анализ в биотехнологических исследованиях	64

Корреляционно-регрессионный анализ в программе Excel	73
Прямолинейная корреляция в программе Excel	73
Прямолинейная регрессия в программе Excel	74
Криволинейная регрессия в программе Excel	78
Работа 8. Дисперсионный анализ данных экспериментов с полной рандомизацией вариантов	85
Дисперсионный анализ данных опытов с полной рандомизацией вариантов в программе Excel	92
Работа 9. Дисперсионный анализ данных однофакторных экспериментов с рандомизированными блоками (повторениями)	95
Дисперсионный анализ данных однофакторных экспериментов с рандомизированными блоками (повторениями) в программе Excel	98
Работа 10. Дисперсионный анализ данных с неоднородными выборками	101
Работа 11. Дисперсионный анализ данных двухфакторного опыта	105
Дисперсионный анализ данных двухфакторного опыта с полной рандомизацией вариантов в программе Excel	110
Работа 12. Планирование биотехнологического опыта (эксперимента)	114
Планирование классического опыта (эксперимента)	114
Планирование регрессионного многофакторного эксперимента	117
Приложения. Статистические таблицы	121
Библиографический список	125

ВВЕДЕНИЕ

Научные исследования в биотехнологии направлены на замену химических продуктов и процессов биологическими, комплексное использование возобновляемого сырья и отходов, расширение спектра целевых продуктов и разработку новых видов материалов, обладающих принципиально новыми свойствами и возможностями, получение новых биоактивных субстанций, повышение эффективности и экологичности сельскохозяйственного производства, обеспечение безопасности и качества продуктов питания. Для приобретения необходимых компетенций в научно-исследовательской деятельности студент должен обладать соответствующими знаниями. В решении этих задач особое место занимает дисциплина «Основы научных исследований в биотехнологии», которая формирует знания и умения по методам биотехнологических исследований, планированию и проведению экспериментов и применению статистических методов оценки экспериментальных данных.

Учебно-практическое пособие составлено в соответствии с рабочей программой дисциплины и предназначено для проведения практических и семинарских занятий по курсу «Основы научных исследований в биотехнологии» по направлению подготовки 19.03.01 «Биотехнология». Тематика разработанных в пособии заданий направлена на формирование у обучающихся общепрофессиональных и профессиональных компетенций.

Пособие включает в себя информационные и индивидуальные задания для самостоятельной работы, методические указания к проведению практических занятий, рабочие таблицы, расчетные формулы, справочный материал и контрольные вопросы для самоподготовки и контроля. Перед решением заданий необходимо проработать рекомендуемую литературу и лекционный материал по теме работы и дать письменные ответы на контрольные вопросы.

При выполнении заданий студентам рекомендуется все работы выполнять по указанным алгоритмам с использованием калькуляторов или в программе Excel, и после статистического анализа сделать биологический вывод.

**Работа 1. Выборочный метод в биологических исследованиях.
Основные статистические показатели данных наблюдений**

Теоретическая часть:

Все методы биологических исследований делятся на *наблюдения и эксперименты* или опыты, в которых изучаемые биологические объекты (растения, микроорганизмы, животные, почва, окружающая среда), во-первых, состоят из большого множества, а во-вторых, они подвержены изменчивости (вариации). В связи с этим, в биотехнологии при проведении наблюдений, испытании и постановке экспериментов пользуются выборочным методом исследований, сущность которого заключается в том, что изучаются не все подлежащие изучению объекты (*генеральная совокупность*), а только часть этих объектов (*выборка*), и по результатам выборки оцениваются, характеризуются все подлежащие изучению объекты – генеральная совокупность. Выборочный метод исследований позволяет сделать оценку (выводы) с определенными допусками (ошибками) и всегда имеет вероятностный характер. Для более точной и объективной оценки генеральной совокупности по результатам выборочного исследования необходимо использовать целый ряд статистических характеристик и критериев. Например, значение генеральной средней (средней генеральной совокупности – μ) при количественной изменчивости можно определить по значению выборочной средней (\bar{x}) и ее ошибки ($S_{\bar{x}}$).

Контрольные вопросы:

1. Наблюдения и эксперимент, их сущность и принципиальное отличие	3. Генеральная совокупность и выборка
2. Основные статистические показатели выборки	4. Значение доверительных интервалов

Общая постановка задачи:

1. Осваивая материал по данной теме, необходимо уяснить: в чем сущность основных методов познания в биологических исследованиях, а именно наблюдение и эксперимент (опыт) и в чем принципиальная разность между этими методами.

2. Познакомиться с простейшим видом научного исследования: *наблюдение* (учеты, измерения, анализы и т.д.).

3. Усвоить основные понятия: совокупность и выборка, признак и его изменчивость, параметры совокупности и статистические характеристики выборки.

Пример1. После применения регуляторов роста на соцветия капусты белокочанной произведен подсчет количества завязей. Получены следующие результаты: 15, 19, 22, 24, 21, 23, 20, 18, 20.

Рассчитать статистические показатели (характеристики) выборки.

Расчет статистических характеристик (для признака X):

Выборочная средняя $\bar{x} = \frac{\sum X}{n} =$

Для расчета суммы квадратов отклонений ($СК = \sum (X - \bar{x})^2$) следует воспользоваться одним из способов:

Таблица 1

Способы определения сумм квадратов отклонений

Значение признака, X	От выборочной средней		От нуля ($A=0$)	
	$(X - \bar{x})$	$(X - \bar{x})^2$	X	X^2
15				
19				
22				
24				
21				
23				
20				
18				
20				
$\sum X =$	$\sum (X - \bar{x}) =$	$\sum (X - \bar{x})^2 =$	$\sum X =$	$\sum X^2 =$

Примечание: для определения суммы квадратов от нуля вводят поправку (корректирующий фактор $-C$):

Поправка $C = \frac{(\sum X)^2}{n} =$

Сумма квадратов (СК) $\sum (X - \bar{x})^2 = \sum X^2 - C =$

Дисперсия $S^2 = \frac{\sum (X - \bar{x})^2}{n-1} =$

Стандартное отклонение $S = \sqrt{S^2} =$

Коэффициент вариации $V = \frac{S}{\bar{x}} \cdot 100 =$

Ошибка выборочной средней $S_{\bar{x}} = \frac{S}{\sqrt{n}} =$

Относительная ошибка средней $E = \frac{S_{\bar{x}}}{\bar{x}} \cdot 100$

95%-й доверительный интервал (ДИ) для генеральной средней (μ):

$$\bar{x} \pm t_{0.5} \cdot S_{\bar{x}} =$$

95%-й доверительный интервал (ДИ) для всей совокупности (любого значения X):

$$\bar{x} \pm t_{0.5} \cdot S =$$

Медиана: $Q_{50} = X_{(n+1)/2}$ (центральное значение ранжированной выборки) при n , равном нечетному числу и $Q_{50} = \left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1} \right) / 2$ (полусумма двух центральных значений) при четном n .

Ранжированный ряд:

Медиана: $Q_{50} =$

Пример 2. Произведен подсчет количества микроорганизмов в молоке в 12 пробах и получены следующие результаты (шт/мл): 5150, 2690, 285, 4750, 1410, 2150, 40500, 890, 1340, 910, 1820, 170

Выборочная средняя $\bar{x} = \frac{\sum X}{n} =$

Медиана: $Q_{50} =$

Квартили:

Нижний Q_{25}

Верхний Q_{75}

Если отношение $n/4$ — дробное число (n — объем выборки), то $Q_{25} = X_{n/4 + 1}$ (из отношения $n/4$ берут только целые). При $n/4$, равном целому числу: $Q_{25} = X_{n/4}$; $Q_{75} = X_{n-n/4}$. Индексы при X : $n/4$, $n/4 + 1$ и $n - n/4$ означают порядковые номера (i) значений X ранжированного ряда.

Медиана: $Q_{50} =$

Представление выборочного распределения в виде «ящика с усами» (Box-plot):

Пример 1:

Пример 2:

Оценка результатов:

Компетенции	Владение теорией	Владение практикой	Балл
Освоены методы познания в биологии. Изучены статистические показатели выборки и параметры генеральной совокупности	+	+	2
Освоено на удовлетворительно	+-	+-	1
Не освоено	-	-	0

Дата _____ Подпись преподавателя _____

Статистическая обработка данных биотехнологических исследований в программе «Excel»

1. Расчеты с использованием МАСТЕРА ФУНКЦИЙ

В программе Excel предусмотрен **Мастер функций**, позволяющий вводить формулы в полуавтоматическом режиме и практически без ошибок. Многие статистические показатели выборки и параметры генеральной совокупности можно очень быстро рассчитать с помощью функций, тем более что наименование большинства функций совпадает с наименованием статистических показателей.

Пример 1. Произведено измерение длины 10-ти колосьев пшеницы ($n=10$), получены следующие результаты (X , см: 6,1; 9,8; 8,2; 8,1; 6,9; 9,1; 9,0; 9,6 8,2; 8,9). Необходимо рассчитать основные статистические показатели выборки.

При запуске программы Excel на экране монитора появляется окно и автоматически создается новая книга (файл) под названием **Книга1**. Документ Excel (файл) имеет расширение ("*.xlsx" в Microsoft Office Excel 2007 и выше) и называется **рабочая книга**. По умолчанию каждая книга состоит из трех рабочих **листов**, в каждый конкретный момент времени активным может быть только один рабочий лист.

После запуска программы Excel в ячейке **B1** Листа **1** пропишем наименование наблюдения «Длина колоса» и ниже в этом столбце внесем значения длины 10 колосьев ($n=10$) (рис.1.1). В ячейки с **A12** по **A17** введем наименование искомым статистических показателей (средняя выборочная, дисперсия, стандартное отклонение и т.д.). Результаты вычислений по указанным показателям будут представлены в ячейках с **B12** по **B17**.

Так как для нашего примера достаточным является представление рассчитанных значений после запятой до сотого знака, необходимо предварительно отформатировать ячейки. Для форматирования ячейки (ячеек) правой кнопкой мыши выделяем их, затем открываем окно диалога **Формат ячеек** или раскрывающийся список **Числовой формат** в группе **Число** на вкладке **Главная** и назначаем необходимый числовой формат.

Для вычисления средней выборочной (средняя арифметическая) активируем ячейку **B12** (**выделенная ячейка со знаком =**), в этой ячейке будут отображаться результаты наших вычислений. Для вызова мастера функций необходимо нажать кнопку **Формулы** на стандартной панели инструментов или в строке формул – на «fx».



Появляется контекстное меню с перечнем различных категорий: **Финансовые**, **Математические**, **Статистические**, **Текстовые**, **Логические** и др.

После выбора категории «**Статистические**» в окне появляется перечень конкретных статистических функций, выбираем **СРЗНАЧ**, что означает среднюю выборочную (среднюю арифметическую) по выборке (рис. 1.1).

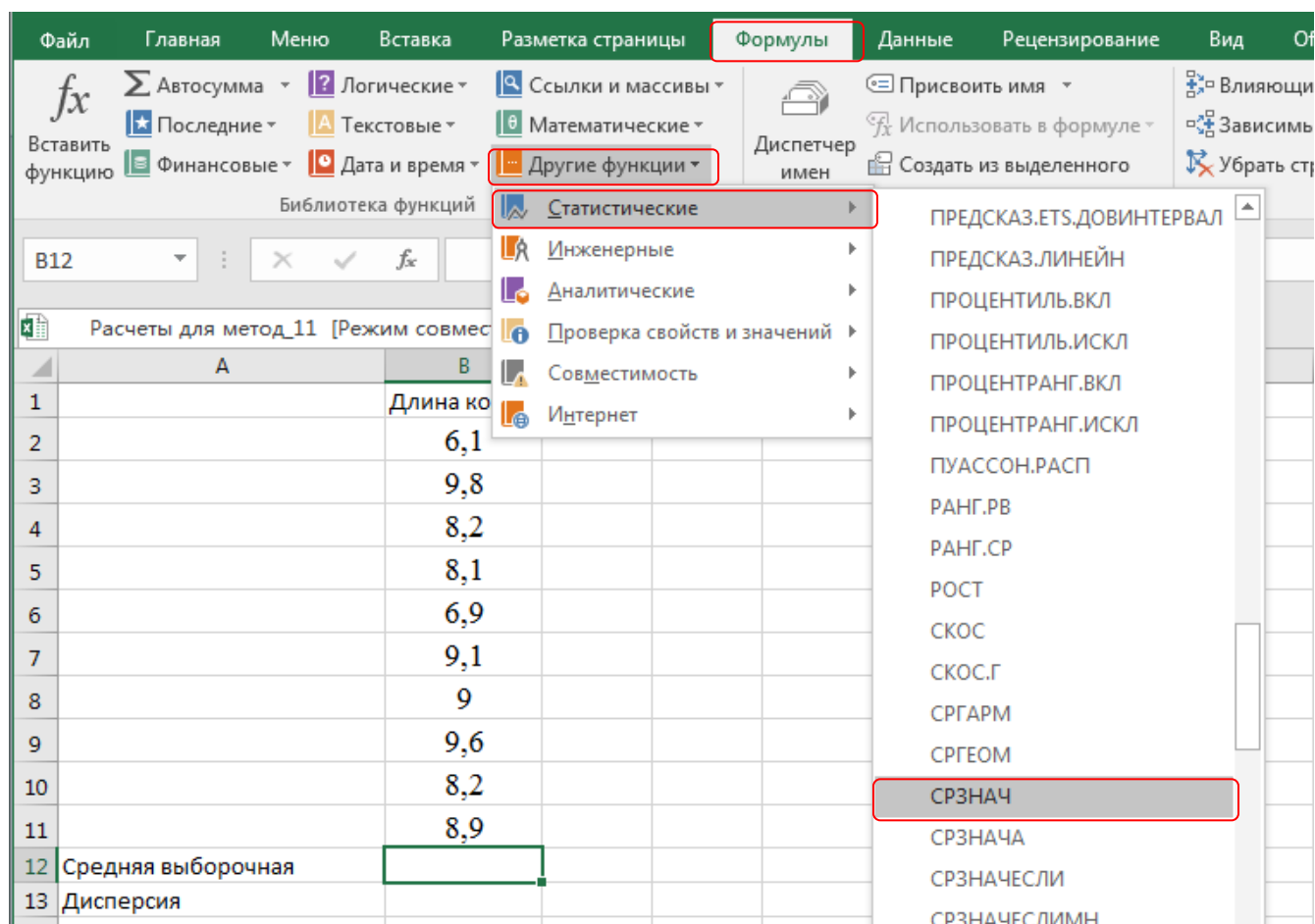


Рис. 1.1. Лист с исходными данными и контекстным меню

Далее открывается окно для выбора аргументов функции (рис. 1.2). В поле **Число 1** ставим курсор и мышкой выбираем диапазон значений длины колоса **B2:B11**, нажимаем клавишу **ОК**, в строке формул автоматически появляется наименование функции и диапазон ячеек (**=СРЗНАЧ(B2:B11)**), а в ячейке **B12** появляется в скобках этот же диапазон (рис. 1.2.)

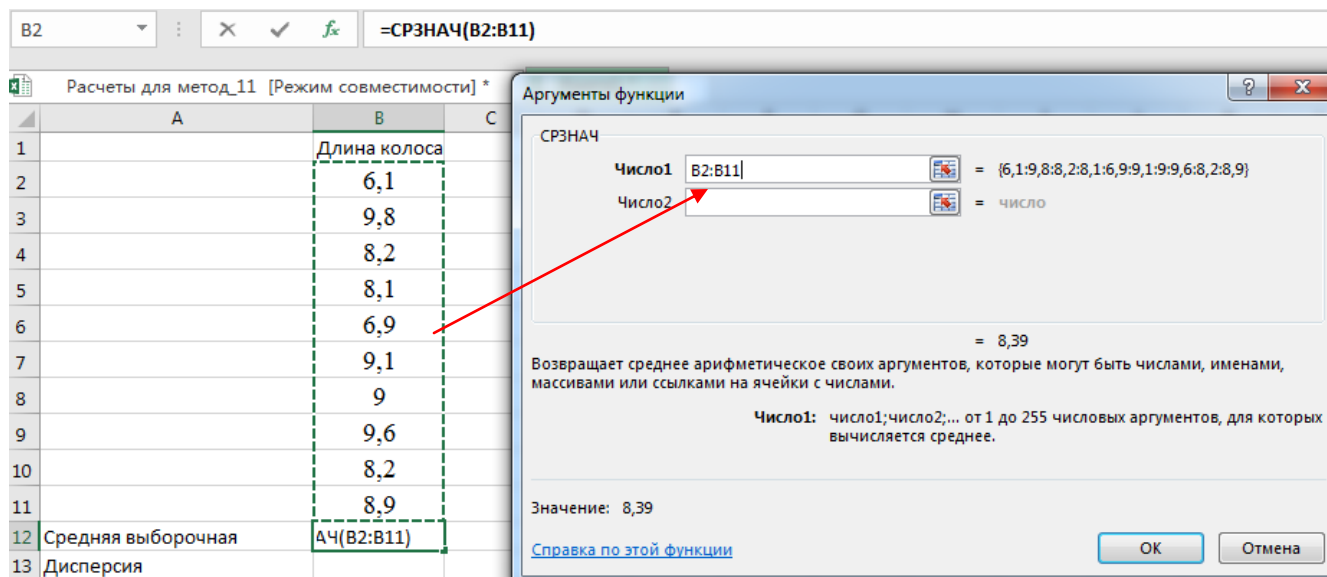


Рис. 1.2. Диалоговое окно для выбора аргументов функции.

После нажатия клавиши **ОК** или щелчка мыши в ячейке **В12** появляется значение средней арифметической (средняя длина колоса) – **8,39 см.** (рис. 1.3.)

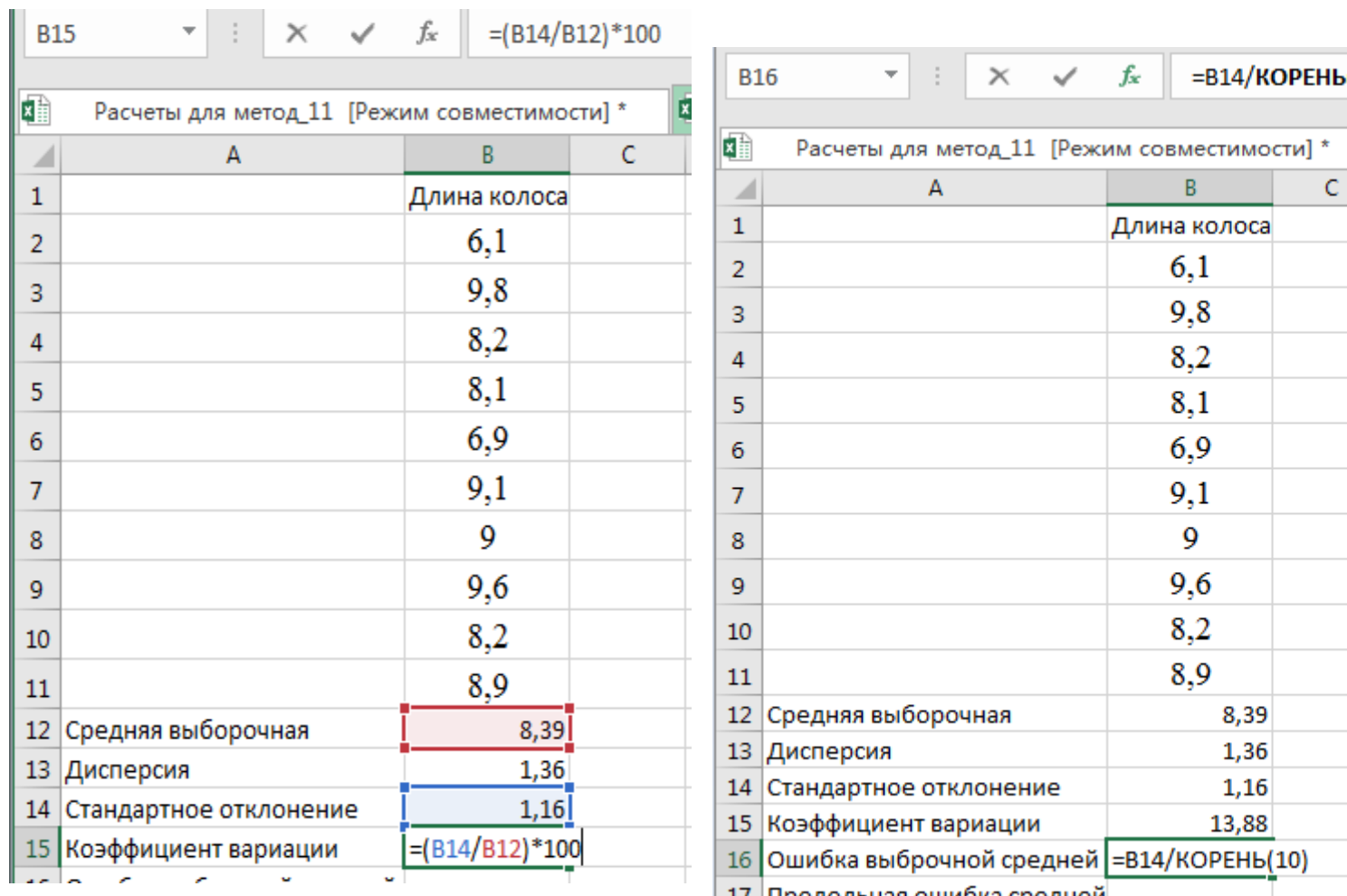


Рис. 1.3. Формулы для определения коэффициента вариации и ошибки выборочной средней

Для расчета дисперсии в MS Excel версии 2010 и выше выбираем из списка функцию **ДИСП.В** – дисперсия для выборки, для тех, кто пользуется MS Excel версия 2007 функцию **ДИСП**. Обратите внимание на то, что в списке имеется функция **ДИСП.Г (ДИСПР)** – дисперсия для генеральной совокупности. Для нашего примера необходимо выбрать **ДИСП.В**, так как рассчитываем дисперсию для выборки. Дисперсия $S^2 = 1,36$ (рис. 1.3). Для нахождения стандартного отклонения выбираем из списка функцию **СТАНДОТКЛОН.В** – стандартное отклонение для выборки. Стандартное отклонение $S = 1,16$ (рис. 1.3).

В списке как математических, так и статистических функций отсутствует функция с помощью, которой можно рассчитать коэффициент вариации (**V**), поэтому воспользуемся процедурой ручного набора формул. Коэффициент вариации представляет собой отношение стандартного отклонения к выборочной средней, выраженной в % ($V = \frac{S}{\bar{x}} \cdot 100$). Выделим ячейку для расчета коэффициента вариации **B15**, затем в строке формул пропишем формулу со ссылкой на ячейки, где находятся стандартное отклонение и выборочная средняя – **(B14/B12)*100** (рис. 1.3) В итоге получаем коэффициент вариации (**V**) = **13,88%**.

Расчет ошибки выборочной средней $S_{\bar{x}} = \frac{S}{\sqrt{n}} =$

Так как в списке функций нет очень важного статистического показателя – ошибки выборочной средней (стандартная ошибка), рассчитаем его с помощью формул. Для этого выделим ячейку для размещения формулы и получения готового результата **B16**, затем в строке формул сначала вставим символ = и пропишем следующую формулу: **B14/КОРЕНЬ (10)**, где в ячейке **B14** – значение стандартного отклонения, 10 – объем выборки (число колосьев). После нажатия на клавишу **Enter** в ячейке **B16** получаем результат $S_{\bar{x}} = 0,37$ (рис. 1.4).

Расчет предельной ошибки выборочной средней. Предельная ошибка выборочной средней представляет собой произведение критерия Стьюдента (*t*) на ошибку выборочной средней ($t_{05} \cdot S_{\bar{x}}$) и служит для нахождения доверительного интервала для генеральной средней.

Для расчета предельной ошибки выборочной средней $t_{05} \cdot S_{\bar{x}}$ в MS Excel версия 2007 выбираем из списка статистических функций функцию **ДОВЕРИТ**.

В Мастере функций в MS Excel версии 2010 и выше для нахождения предельной ошибки средней выборочной имеется функция, которая имеет название **ДОВЕРИТ. СТЬЮДЕНТ**. Поместим курсор в ячейку **B17**. затем из списка статистических функций выберем функцию **ДОВЕРИТ. СТЬДЕНТ** (рис. 1.4.), нажимаем **ОК**.

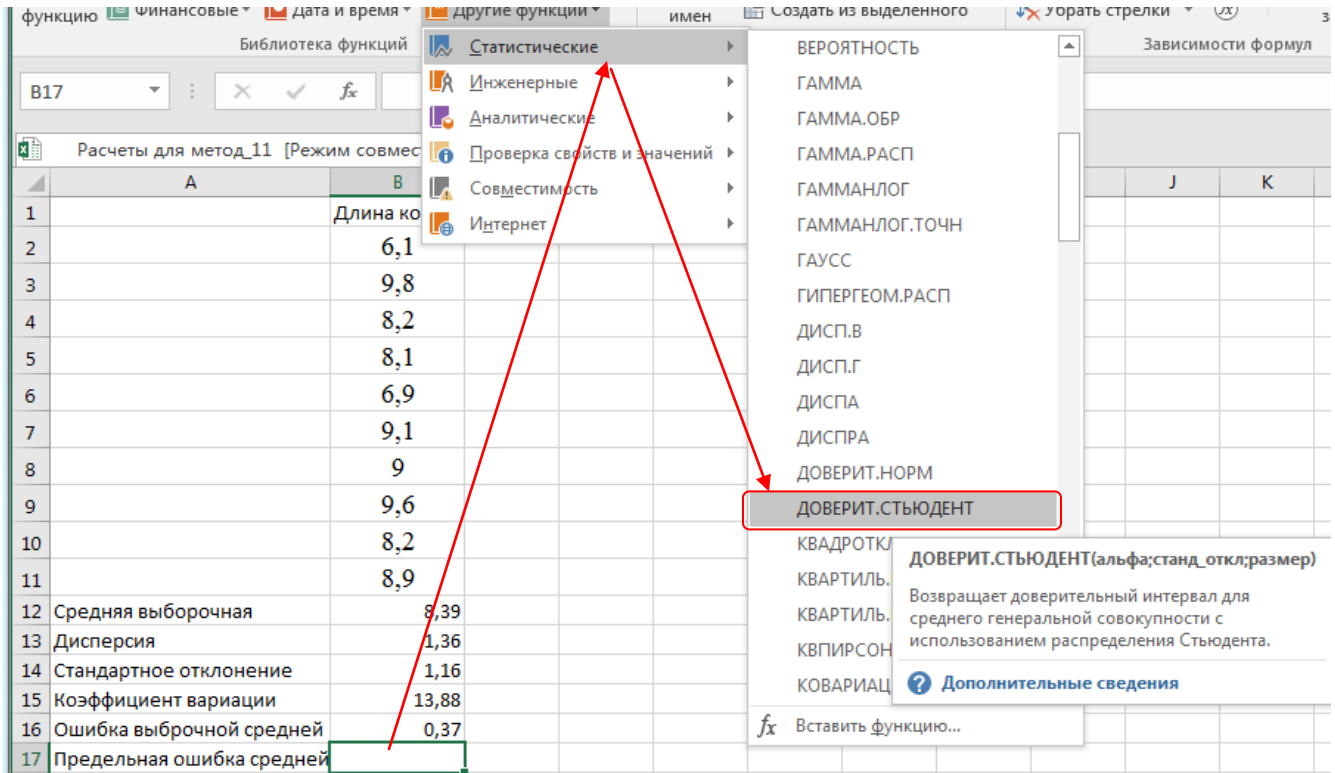


Рис. 1.4. Расчет предельной ошибки выборочной средней

В появившемся диалоговом окне вводим:

- в поле **Альфа** уровень значимости – **0,05**,
- в поле **Станд_откл.** – ссылку на ячейку (**В14**), где находится стандартное отклонение или готовое значение (**1,164712**)
- в поле **Размер** – объем выборки (**10**)

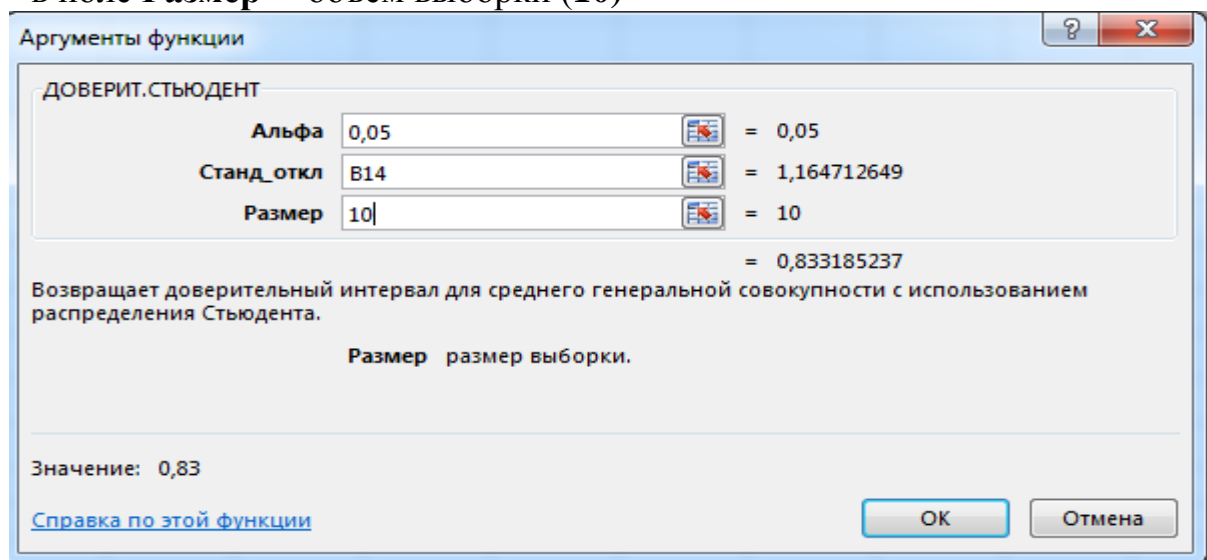


Рис. 1.5. Выбор в меню функции ДОВЕРИТ. СТЬЮДЕНТ (предельная ошибка средней)

После нажатия на клавишу **ОК** получаем результат – предельная ошибка выборочной средней равна **0,83** (рис. 1.6.)

B17 f_x =ДОВЕРИТ.СТЪЮДЕНТ(0,05;B14;10)

Расчеты для метод_11 [Режим совместимости] * Книга1 * x

	A	B	C	D	E
1		Длина колоса			
2		6,1			
3		9,8			
4		8,2			
5		8,1			
6		6,9			
7		9,1			
8		9			
9		9,6			
10		8,2			
11		8,9			
12	Средняя выборочная	8,39			
13	Дисперсия	1,36			
14	Стандартное отклонение	1,16			
15	Коэффициент вариации	13,88			
16	Ошибка выборочной средней	0,37			
17	Предельная ошибка средней	0,83			

Рис. 1.6. Результаты расчетов основных статистических показателей выборки

95%-й доверительный интервал (ДИ) для генеральной средней (μ):

$$\bar{x} \pm t_{0,5} \cdot S_{\bar{x}} = 8,39 \pm 0,83$$

2. Расчеты с использованием надстройки Анализ данных

В MS Excel встроен в качестве надстройки пакет **Анализ данных**, который в значительной степени ускоряет и облегчает расчеты по статистической обработке результатов исследований, что дает возможность ввести исходные данные и в автоматическом режиме сразу получить результаты, проверить гипотезу по двум выборкам, провести дисперсионный, корреляционный и регрессионный анализы.

Далее все расчеты проведем с использованием **Анализа данных**. Для выполнения различных статистических задач в надстройке **Анализ данных** имеются 19 инструментов: Описательная статистика, Гистограмма, Корреляция, Регрессия и др.

Для выполнения расчетов, прежде всего, необходимо проверить, настроена ли Ваша программа Excel на использование надстройки **Анализ данных**. Если при запуске Excel во вкладке **Данные** не отображается кнопка **Анализ данных**, то следует подключить надстройку **Анализ данных MS Excel**.

Подключение надстройки Анализ данных в MS Excel версии 2010 и выше

Откройте вкладку **Файл**, нажмите кнопку **Параметры** и выберите категорию **Надстройки**. В раскрывающемся списке **Управление** выберите пункт **Надстройки Excel** и нажмите кнопку **Перейти** (рис. 1.7)

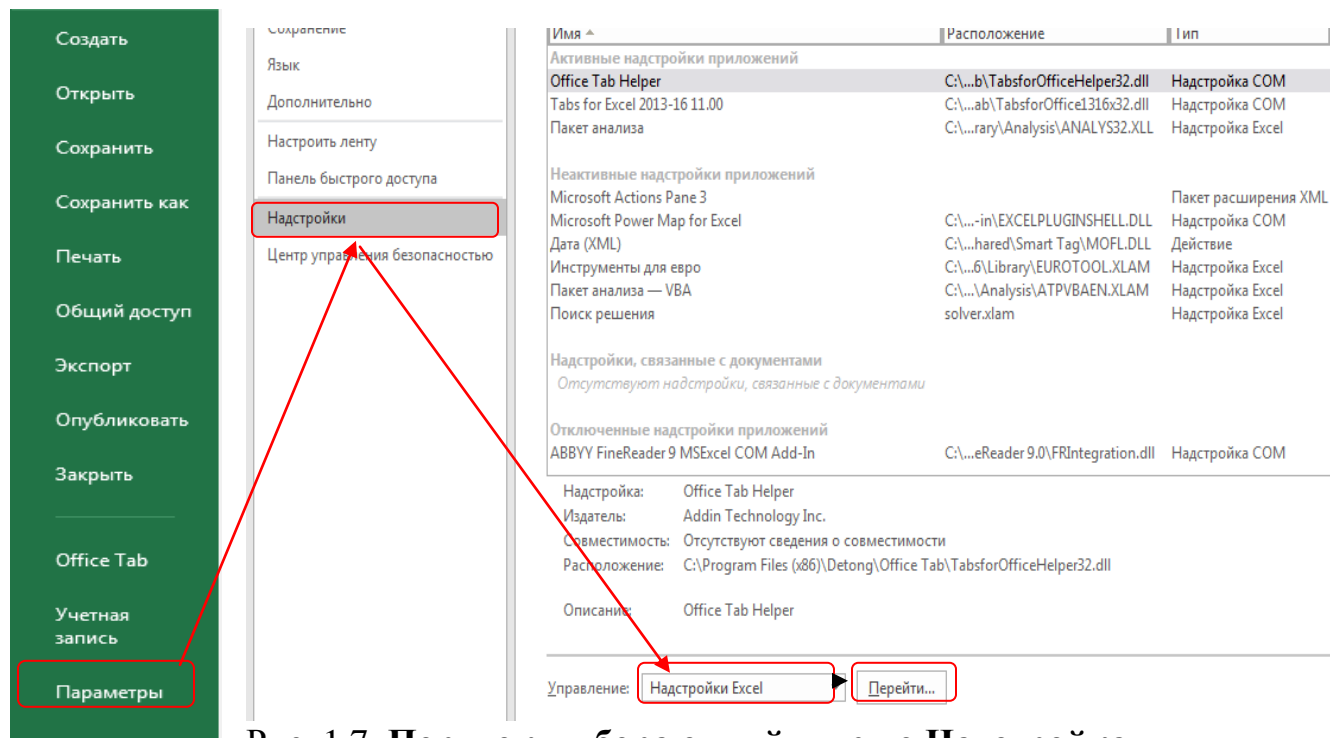


Рис. 1.7. Порядок выбора опций в меню **Надстройка**

В диалоговом окне **Надстройки** установите флажок **Пакет анализа**, а затем нажмите кнопку **ОК** (рис.1.8). Если **Пакет анализа** отсутствует в списке поля **Доступные надстройки**, нажмите кнопку **Обзор**, чтобы выполнить поиск. Если выводится сообщение о том, что пакет анализа не установлен на компьютере, нажмите кнопку **Да**, чтобы установить его.

После нажатия кнопки **Пакет** появляется на ленте функций. Для получения доступа к нему в интерфейсе программы выбирается вкладка **Данные**, в правой части меню **Раздел анализа** появляется иконка опции **Анализ данных**.

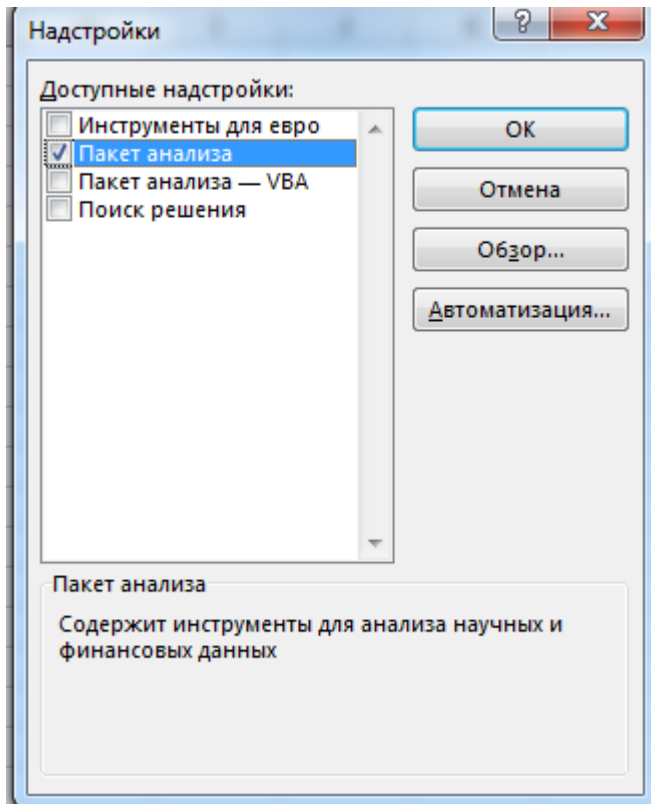



Рис.1.8. Выбор в надстройках Пакета анализа

Подключение надстройки Пакет анализа в MS Excel версия 2007:

Подвести курсор к значку «Кнопка «Office»», , расположенному в верхнем левом углу окна Excel, нажать на правую клавишу мышки, затем выбрать «Настройка панели быстрого доступа», (рис. 1.9) и нажать Ввод.

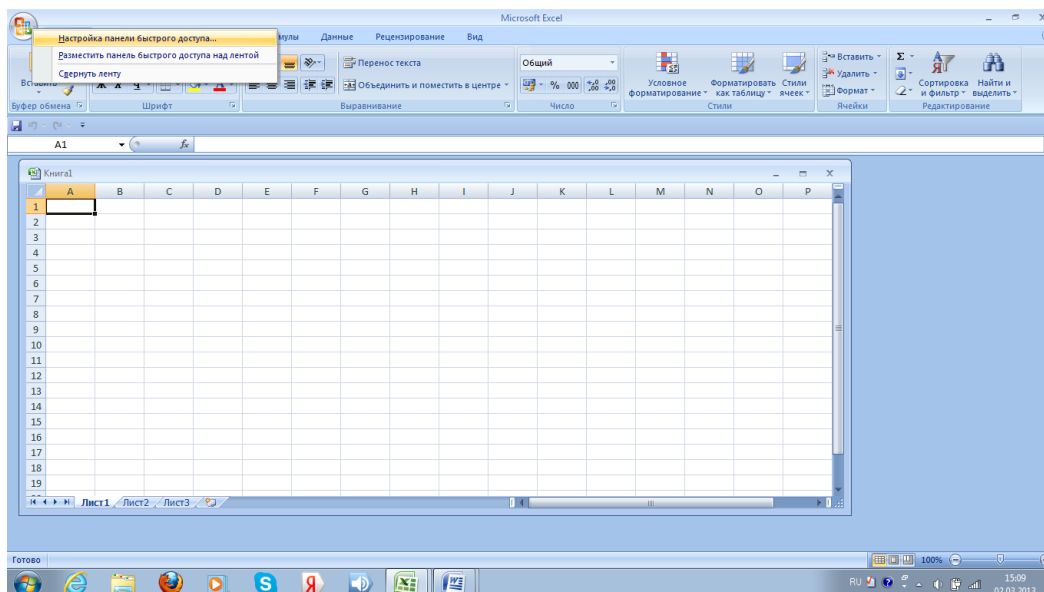


Рис. 1.9. Окно программы Excel

После ввода открывается окно **Параметры**, выбрать **Надстройки**, нажать на вкладку «**Перейти**», в появившемся окне **Надстройки** установите флажок **Пакет анализа** и нажать **ОК**. и тем самым активируете **Пакет анализа** (рис.1.10)

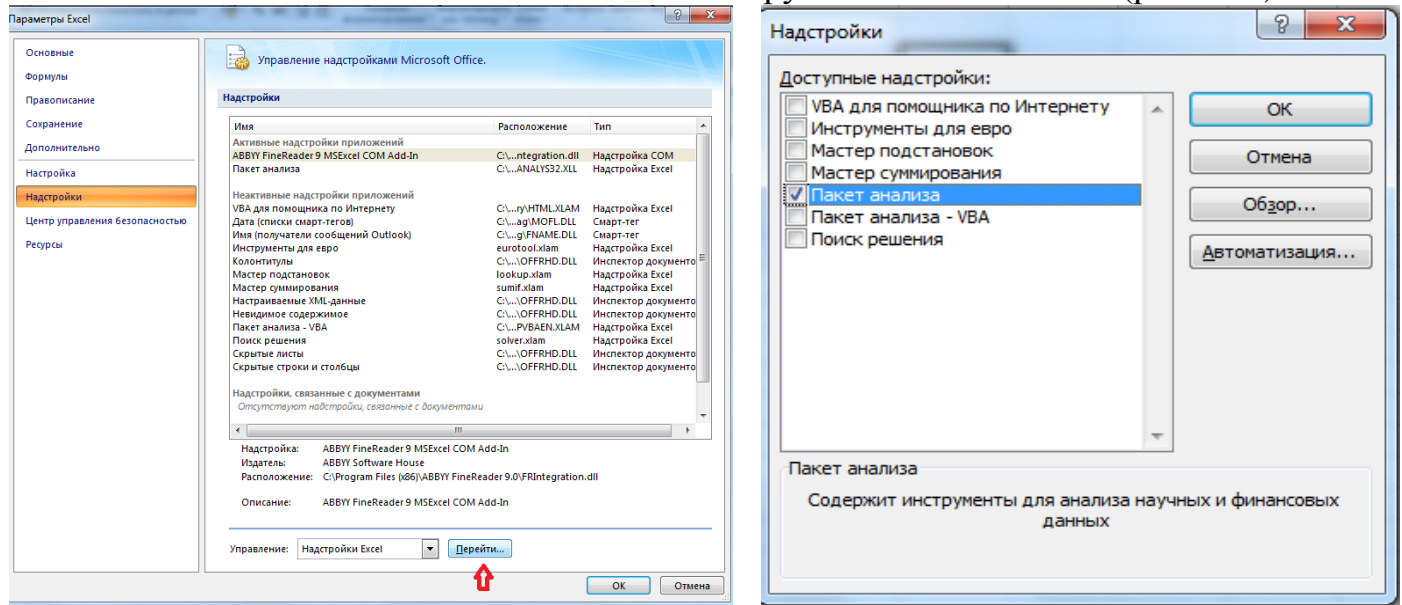


Рис. 1.10. Активирование надстройки **Пакет анализа** в MS Excel -2007

3. Расчет основных статистических показателей выборки с использованием **Описательной статистики Пакета анализа**

В активный лист программы Excel ввести в столбце **В** наименование признака «**Длина колоса**» и результаты измерения по 10 колосьям:
 Во вкладке **Данные** нажать на клавишу **Анализ данных**, выбрать инструмент анализа **Описательная статистика** (рис.1.11, рис. 1.12.)

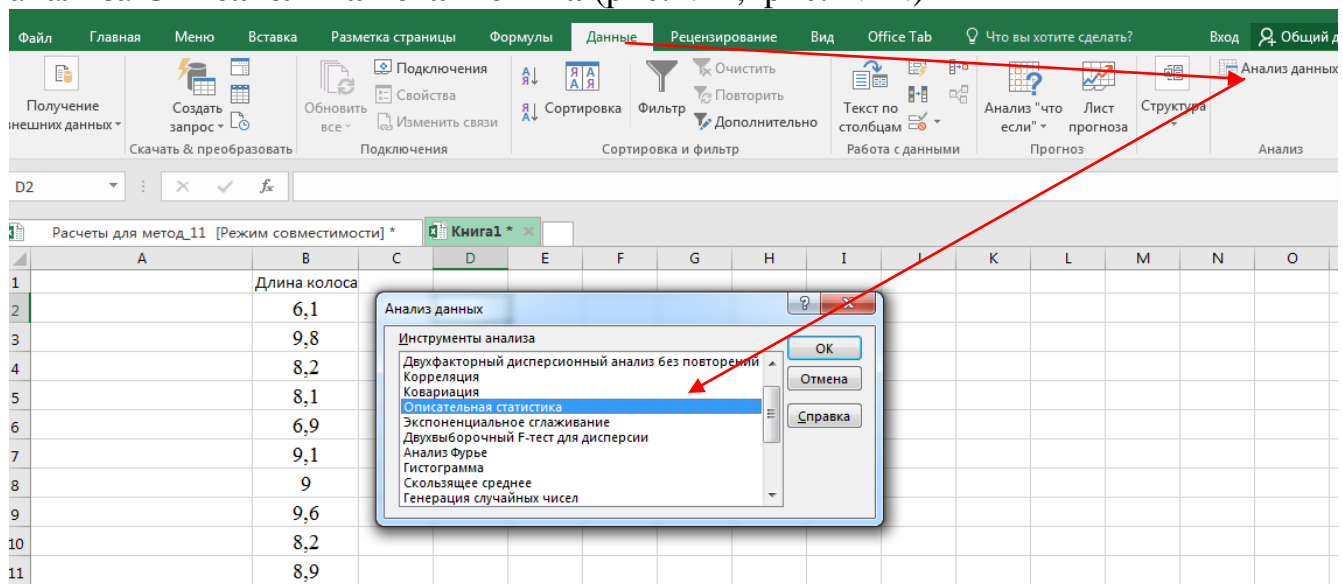


Рис.1.11. Исходные данные и выбор инструмента «**Описательная статистика**»

В окне **Описательная статистика** в поле **Входной интервал:** ввести с помощью мышки диапазон ячеек **B1:B9** и нажать **Ввод**, выбрать **Группирование по столбцам**, поставить галочку в поле **Метки в первой строке** (в этом случае в заголовке результатов будет автоматически прописано «Длина колоса»), в поле **Выходной интервал** можно выбрать размещение на рабочем листе с исходными данными, **Новый рабочий лист** или **Новая рабочая книга**.

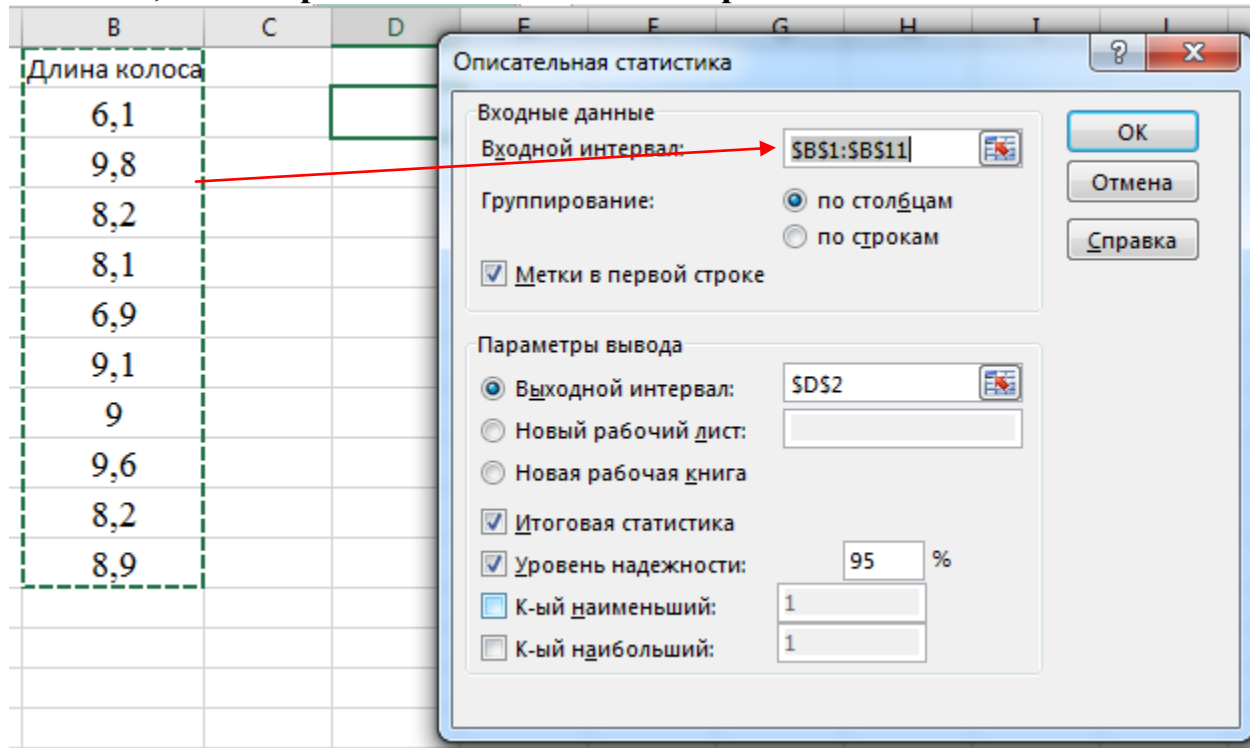


Рис. 1.12. Выбор параметров в окне *Описательная статистика*

В случае размещения на листе 1 рядом с исходными данными в поле **Выходной интервал** необходимо указать интервал ячеек для размещения итоговых результатов описательной статистики, в нашем примере интервал **D2**. Далее выбираем **Итоговую статистику**, в поле **Уровень надежности** указываем **95**, в полях **К-тый наименьший** и **К-тый наибольший** ставим **1** или игнорируем, нажимаем **ОК** (рис. 1.12.).

После нажатия на клавишу **ОК** получаем результаты по основным статистическим показателям выборки (рис. 1.13), которые можно было рассчитать с помощью ввода формул или выбора соответствующих функций для каждого показателя выборки, что заняло бы достаточно много времени и усилий для большого числа операций.

В	С	Д	Е	Ф
Длина колоса				
6,1		Длина колоса		
9,8				
8,2		Среднее	8,39	
8,1		Стандартная ошибка	0,37	
6,9		Медиана	8,55	
9,1		Мода	8,20	
9		Стандартное отклонение	1,16	
9,6		Дисперсия выборки	1,36	
8,2		Эксцесс	0,29	
8,9		Асимметричность	-0,88	
		Интервал	3,70	
		Минимум	6,10	
		Максимум	9,80	
		Сумма	83,90	
		Счет	10,00	
		Уровень надежности(95,0%)	0,83	

Рис. 1.13. Результаты расчета основных статистических показателей выборки

На рис. 1.14. представлены итоговые данные по работе 1 рассчитанные с помощью статистических функций (слева) и с помощью **Пакета анализа** (справа).

Выбор в пользу **Пакета анализа** при статистических расчетах по сравнению с двумя вышеприведенными способами вполне очевиден, так как после ввода исходных данных двумя-тремя кликами мышки мы получаем готовый результат.

Недостатком **Пакета анализа** является то, что в **Описательной статистике** нет таких статистических показателей, как «Ошибка выборочной средней», «Коэффициент вариации», «Доверительный интервал для генеральной средней» и «Доверительный интервал для всей совокупности». Однако указанные показатели можно рассчитать с помощью формул в программе Excel или на калькуляторе.

А	В	С	Д	Е	Ф
	Длина колоса				
1			Длина колоса		
2		6,1			
3		9,8			
4		8,2			
5		8,1			
6		6,9			
7		9,1			
8		9			
9		9,6			
10		8,2			
11		8,9			
12	Средняя выборочная	8,39			
13	Дисперсия	1,36			
14	Стандартное отклонение	1,16			
15	Коэффициент вариации	13,88			
16	Ошибка выборочной средней	0,37			
17	Предельная ошибка средней	0,83			

Рис.1.14. Итоговые данные на основе статистических функций (слева) и Пакета данных (справа)

Работа 2. Группировка и графическое представление данных наблюдений

Теоретическая часть: Наблюдения над биологическими объектами в опытах, как правило, проводятся одновременно по одному или нескольким признакам, что позволяет собрать наиболее полные сведения о влиянии факторов на изучаемые биотехнологические процессы. Обширный цифровой материал нуждается в обработке, которая начинается с упорядочения собранных данных, с тем, чтобы извлечь заключенную в них информацию. Процесс систематизации или упорядочивания данных наблюдений и учетов с целью расчета статистических показателей и построения кривой вариационного ряда называется *группировкой*.

Для наглядного выражения закономерностей варьирования того или иного количественного признака вариационные ряды изображают в виде геометрических фигур в системе прямоугольных координат и получают при непрерывной изменчивости гистограмму, а дискретной – полигон.

Различают *эмпирические и теоретические распределения* частот совокупности результатов наблюдений. Эмпирическое распределение – это распределение результатов наблюдений, полученных при изучении выборки. В основе его лежат определенные математические закономерности, которые в генеральной совокупности, т.е. при большом числе наблюдений ($n \rightarrow \infty$) характеризуются некоторыми теоретическими распределениями. Наиболее часто в биологических исследованиях используются следующие теоретические распределения: нормальное, биномиальное, распределение Пуассона, а также специальные (*t-распределение Стьюдента, F-распределение Фишера, χ^2 -распределение Пирсона* и др.). На основе теоретических распределений построены статистические критерии, которые используются при проверке гипотез.

Результаты различных наблюдений и экспериментов чаще всего располагаются приблизительно в соответствии с симметричной кривой нормального распределения. Но нередко некоторые признаки растений, насекомых и почвы значительно отличаются от нормального – ассиметричные, скошенные, многовершинные кривые и т.д.

Контрольные вопросы к защите:

1. Виды изменчивости. Порядок группировки данных выборки.	2. Закономерности нормального распределения. Особенности распределения Стьюдента
3. Расчет (планирование) объема выборки. Репрезентативность выборки.	4. Причины появления ассиметричных кривых.

Общая постановка задачи:

1. Научиться группировать данные при большом числе исходных данных.
2. Изобразить графически значения вариационного ряда.
3. Рассчитать статистические показатели (характеристики) данных выборки.

Построить гистограмму и кривую вариационного ряда.

4. На основании анализа графического изображения определить общие закономерности о соответствии эмпирического ряда кривой нормального распределения.

Список индивидуальных заданий:

После получения номера примера и задания от преподавателя, выпишите соответствующие колонки цифр в рабочую таблицу, проведите группировку и статистическую обработку, представьте гистограмму распределения значений выборки и сделайте вывод.

Таблица 2

Шифр задания к работе 2

№ задания	Номера колонок	№ задания	Номера колонок	№ задания	Номера колонок
1	1 2 3 4 5	11	5 7 8 9 10	21	1 2 3 4 5 6
2	2 3 4 5 6 8	12	1 4 5 6 7 9	22	1 2 3 4 7
3	3 4 5 6 7	13	2 5 6 7 8	23	1 2 3 6 8
4	4 5 6 8 9	14	3 6 7 8 9	24	1 2 3 4 9
5	5 6 7 8 9	15	4 7 8 9 10	25	1 2 3 4 10
6	1 3 6 7 8 9	16	1 5 6 7 8	26	2 3 4 5 7 9
7	1 3 4 5 6	17	2 6 7 8 9	27	3 5 7 8 9
8	2 4 5 6 7	18	3 7 8 9 10	28	5 6 8 9 10
9	2 3 5 6 7 8	19	1 2 6 7 8 9	29	1 3 4 5 7 9
10	4 6 7 8 9	20	2 7 8 9 10	30	4 5 7 8 9

Таблица 3

Пример. Длина колоса озимой пшеницы, см

1	2	3	4	5	6	7	8	9	10
6,0	7,1	7,7	8,3	6,9	7,9	6,1	8,4	6,6	6,2
7,3	6,7	7,4	8,4	7,5	7,4	9,8	8,1	6,9	8,4
6,9	7,2	6,9	8,0	7,4	7,8	8,2	9,0	6,7	8,4
6,5	7,2	8,4	7,5	8,0	7,1	8,1	7,9	7,8	7,6
8,3	7,8	8,5	8,1	8,4	7,5	6,9	7,4	6,9	7,0
7,7	7,8	7,0	7,6	6,9	8,6	9,1	6,0	8,0	10,2
8,8	6,9	7,1	7,5	6,9	7,4	9,0	7,8	7,5	7,4
6,3	10,0	6,9	7,8	7,0	7,7	9,6	9,0	6,4	9,0
6,3	7,8	7,0	9,2	6,8	8,0	8,2	9,3	7,8	7,5
8,1	8,5	8,3	7,6	9,4	–	8,9	9,0	–	7,5

Таблица 4

Рабочая таблица

Номера колонок	Значения признака, X

Порядок группировки:

1. Определите число классов (групп) $k =$
2. Рассчитайте классовый интервал $i = \frac{X_{\max} - X_{\min}}{k} = \frac{R}{k} =$
3. Установите для каждой группы (класса) нижние и верхние границы
4. Проведите статистическую обработку на основе групповых средних (табл. 5):

Таблица 5

Расчетная таблица ($A=$)

Группа (класс)	Частота f	Среднегрупповая, M	$x_l^* = M - A$	$f \cdot x_l$	x_l^2	$f \cdot x_l^2$
Суммы:	$\sum f = n =$	—	$\sum f \cdot x_l =$		$\sum f \cdot x_l^2 =$	

^{*} x_l -отклонение от среднегрупповой

$$\bar{x} = A + (\sum f \cdot x_l) : n =$$

$$CK = \sum (X - \bar{x})^2 = \sum f \cdot x_l^2 - (\sum f \cdot x_l)^2 : n =$$

$$S^2 = \frac{\sum (X - \bar{x})^2}{n - 1} =$$

$$S = \sqrt{S^2} =$$

$$V = \frac{S}{\bar{x}} \cdot 100 =$$

$$S_x = \frac{S}{\sqrt{n}} =$$

$$\bar{x} \pm t_{0.05} \cdot S_x =$$

$$\bar{x} \pm t_{0.05} \cdot S =$$



Рис.1. Гистограмма и кривая эмпирического распределения
 _____ (значений изучаемого признака)

Заключение:

Оценка результатов:

Компетенции	Владение теорией	Владение практикой	Балл
Владеет методом группировки данных большой выборки и анализа данных наблюдений	+	+	2
Владеет на удовлетворительно	+-	+-	1
Не владеет	-	-	0

Дата _____ Подпись преподавателя _____

Группировка данных, расчет статических показателей, построение гистограммы в программе Excel

Пример. Произведено взвешивание 50 клубней ($n = 50$) картофеля, г.
70 64 135 76 83 12 75 60 145 85 11 112 68 56 42 109 94 67 98 75 150 86 49 35 120 83
89 125 76 45 9 34 85 75 63 102 99 83 65 150 35 65 35 84 134 142 76 145 59 70
Необходимо рассчитать основные статистические показатели выборки, построить гистограмму.

1. В активный лист программы Excel ввести в ячейке **B1** наименование статистического приема «Группировка», в ячейке **B2** – Масса клубня, г, а в диапазоне с **B3** по **B52** значения массы 50 клубней картофеля.

2. Во вкладке **Данные** нажать на кнопку **Анализ данных** и выбрать инструмент анализа **Гистограмма** (рис.2.1.)

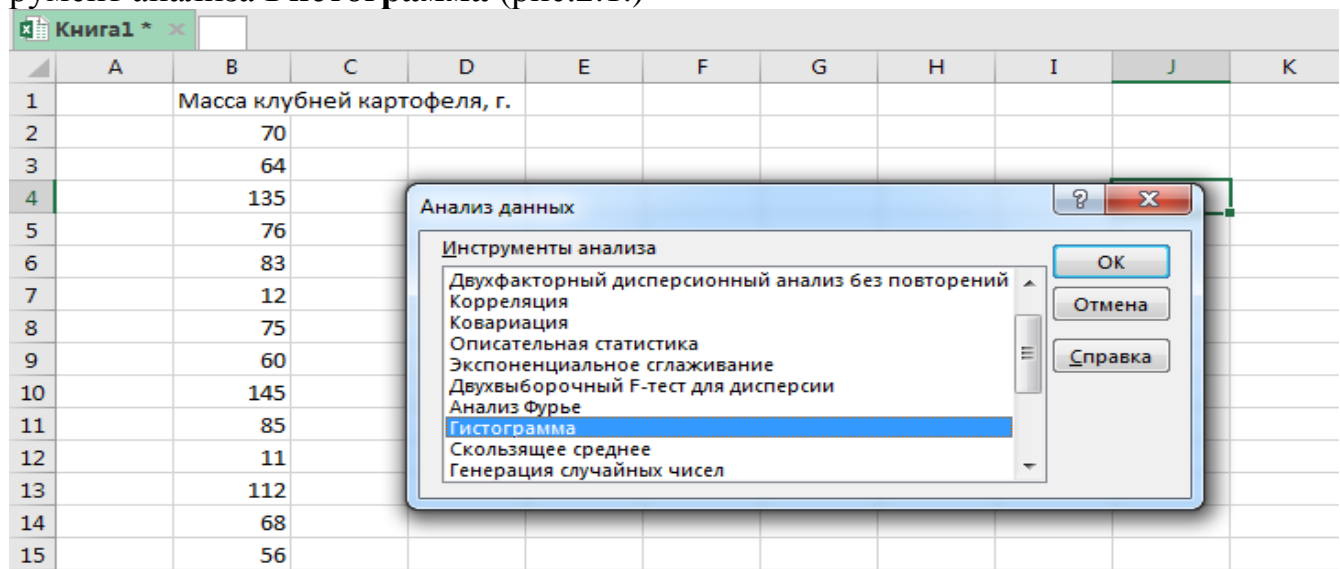


Рис. 2.1. Рабочая книга с исходными данными и выбор инструмента «Гистограмма»

3. В окне **Гистограмма** в поле **Входной интервал:** ввести с помощью мышки диапазон ячеек **B2:B52** и нажать **Ввод**, в поле **Интервал карманов**, что представляет собой классовый интервал, можно указать известный интервал или как в нашем случае, оставить его пустым. Поставить галочку в поле **Метки** (в этом случае в заголовке результатов будет автоматически прописано «Масса клубня, г»), в поле **Выходной интервал** можно выбрать размещение на рабочем листе с исходными данными или **Новый рабочий лист**.

4. В нашем примере гистограмма будет размещаться на первом рабочем листе с исходными данными, в поле **Выходной интервал** укажем любой диапазон ячеек, например, **E7:H7**, далее выбираем **Вывод графика**. После заполнения всех полей нажимаем **ОК** (рис. 2.2).

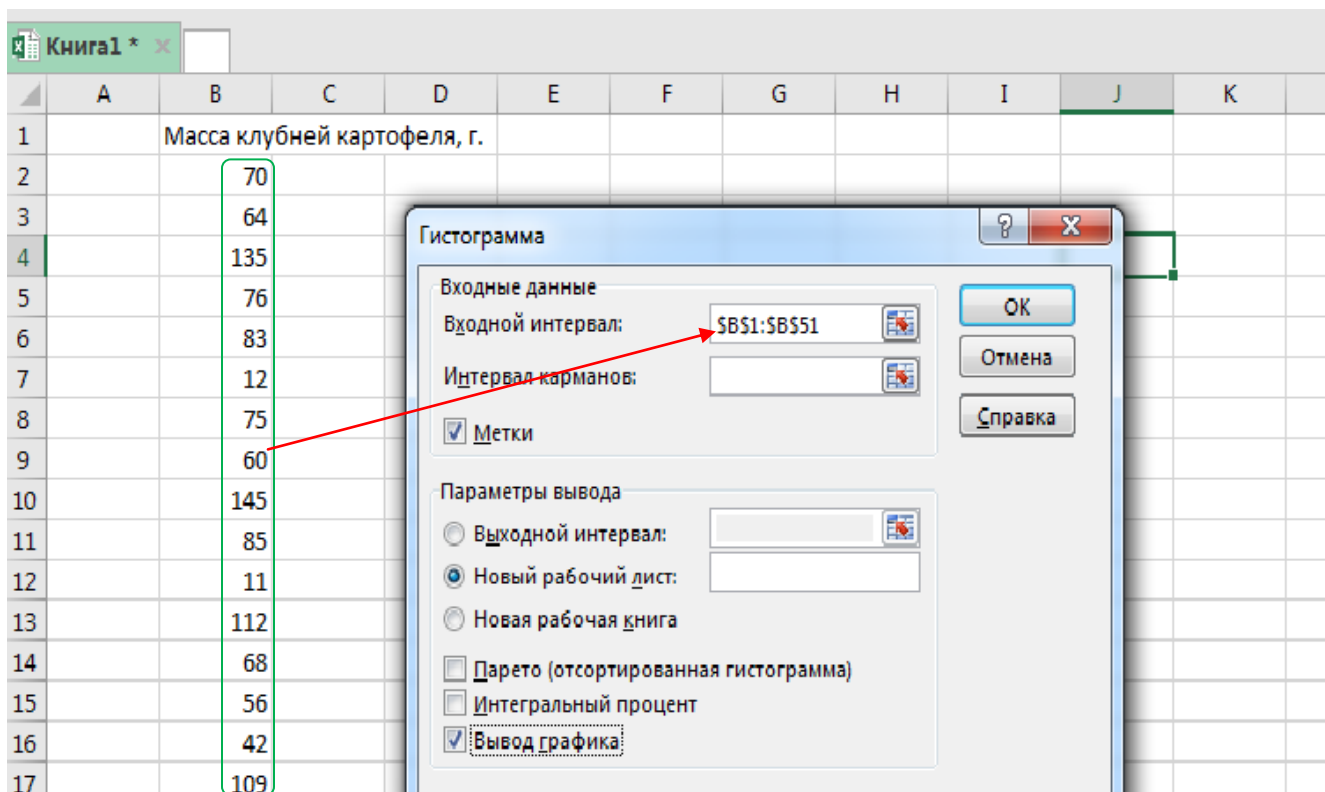


Рис. 2.2. Выбор входных данных и установление параметров вывода

5. После нажатия на клавишу **OK** получаем рядом с исходными данными, таблицу вариационного ряда и гистограмму распределения 50 клубней картофеля по массе (рис. 2.3.).

Здесь под термином «карман» подразумевается общепринятый в статистике термин «группы» или «классы». В таблице помимо абсолютной численности частот можно представить относительные частоты, причем накопленные или так называемые интегральные.

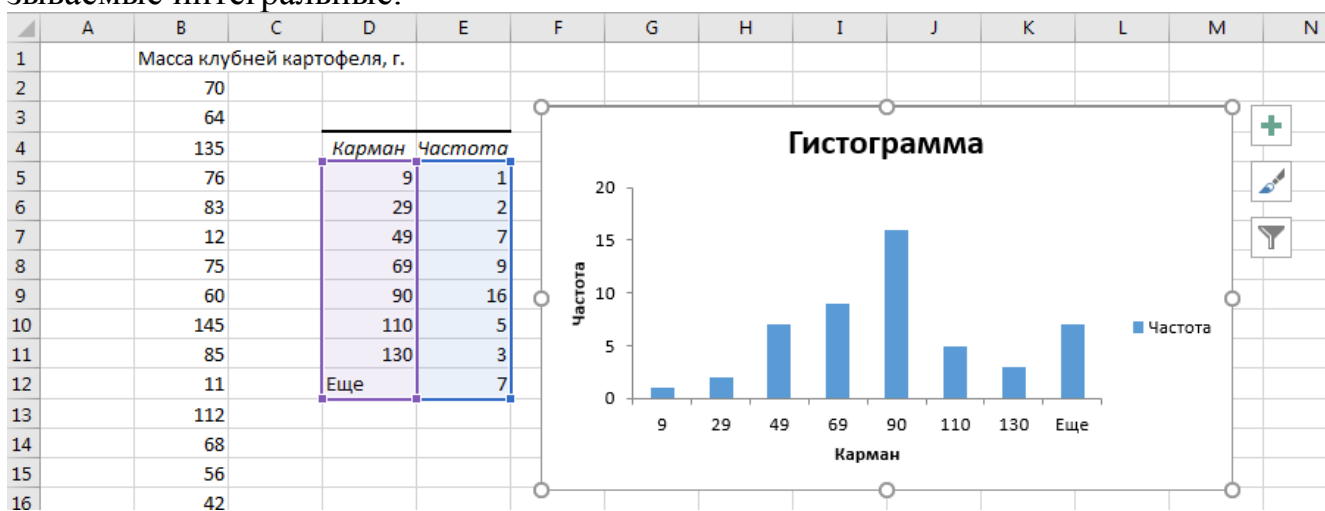


Рис.2.3. Таблица вариационного ряда и гистограмма распределения 50 клубней картофеля

6. Так как гистограмма выглядит в нашем окне нечитаемой и непрезентабельной, следует ее растянуть по вертикали и горизонтали на этом Листе или же скопировать ее и перенести на Лист 2.

7. Для лучшей наглядности наведем мышкой курсор на гистограмму, как только рисунок станет активным, на что указывает подсказка **Область построения**, можно будет растянуть гистограмму так, чтобы на рисунке были видны все частоты и значения по классам или группам (карман).

8. Для удаления больших зазоров между рядами (столбиками) гистограммы и корректировки параметров рядов необходимо подвести курсор мышки к любому ряду и нажать на правую клавишу мышки и в выпадающей панели выбрать **Формат ряда данных** (рис.2.4).

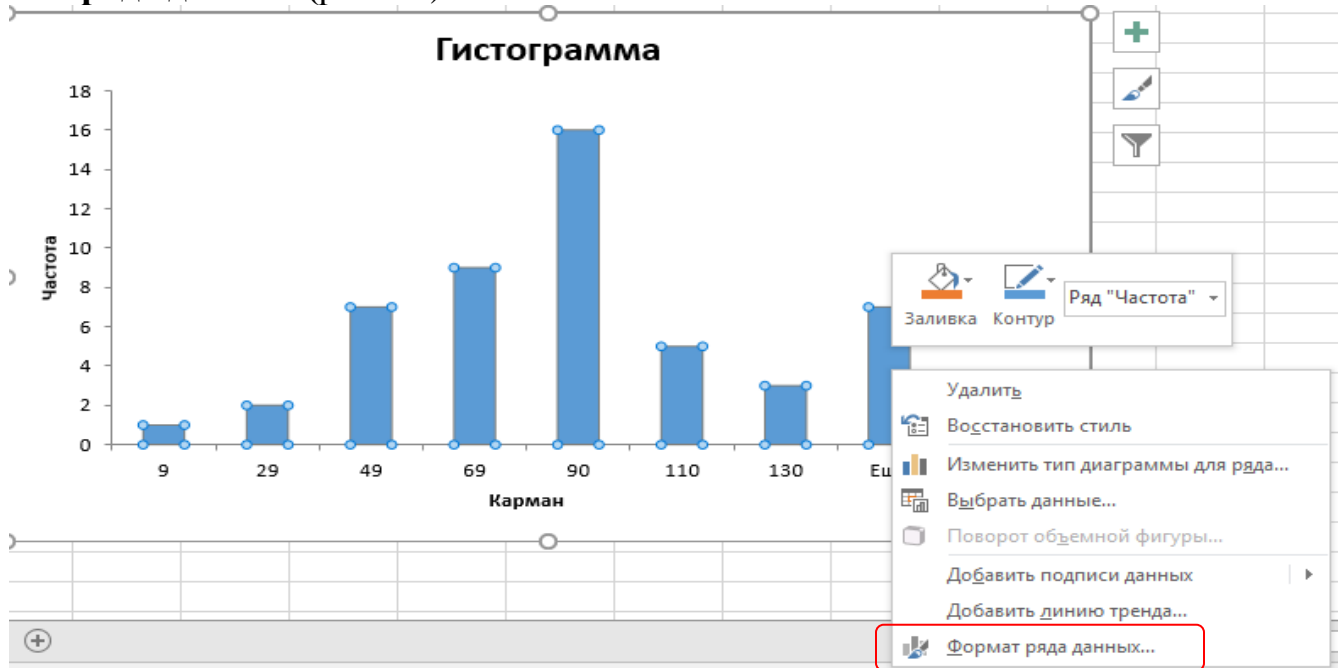


Рис. 2.4. Выбор формата ряда данных

8. После нажатия на клавишу **Ok** появляется окно **Формат ряда данных**, в котором можно изменить параметры одного или всех рядов. Прежде всего, устраняем боковые зазоры, для чего передвигаем в крайнее левое положение бегунок бокового зазора, который по умолчанию занимает среднее положение (рис. 2.5.). Кроме того, с помощью заливки можно выбрать необходимый цвет рядов, выделить их границы.

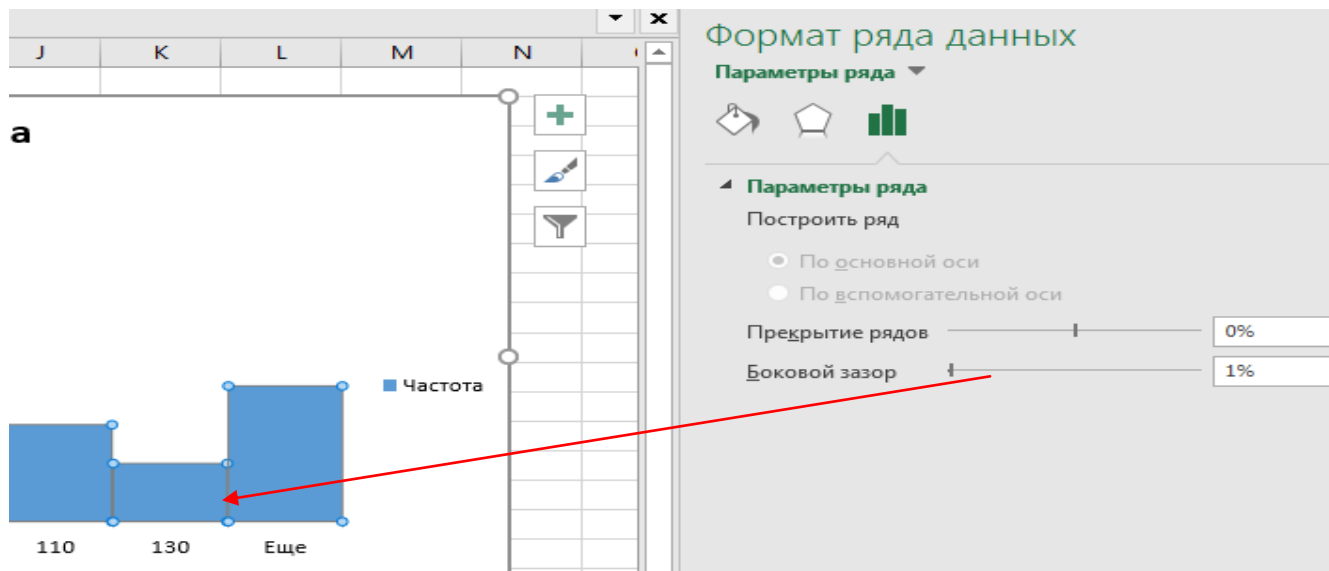


Рис. 2.5. Форматирование параметров ряда

9. В итоге получаем следующую Гистограмму распределения массы 50 клубней картофеля (рис. 2.6.)

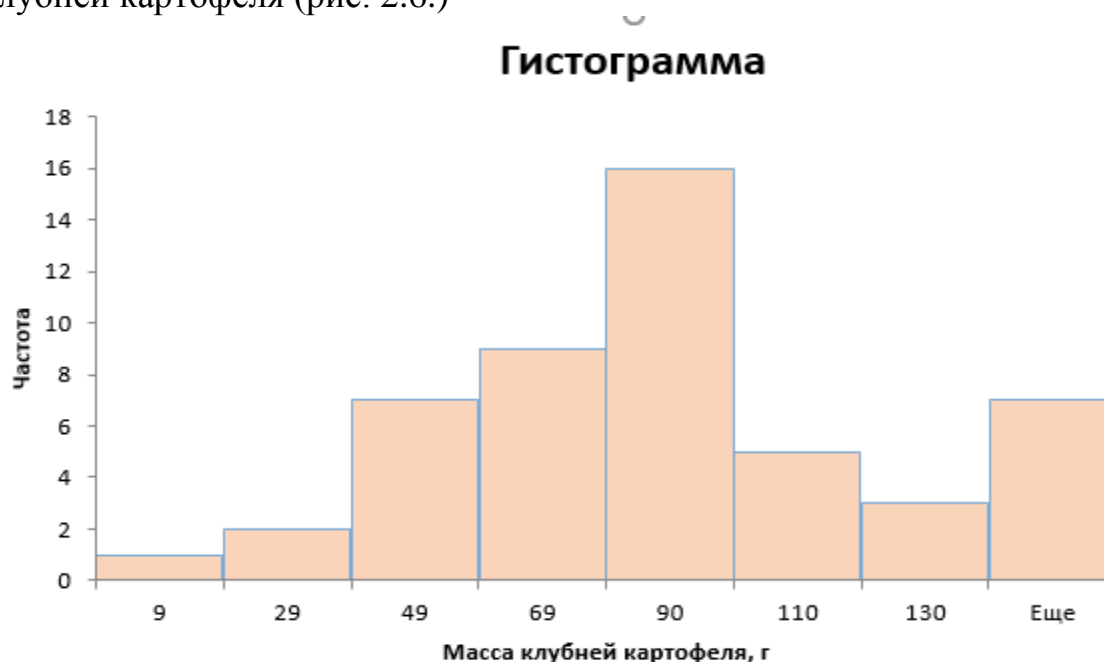


Рис. 2.6. Гистограмма распределения массы 50 клубней картофеля

С введенными на Листе исходными данными можно провести и другие статистические операции: рассчитать статистические показатели, проверить на нормальность данных и др.

Проведем расчеты основных статистических показателей с использованием в **Анализе данных** **Описательной статистики**.

1. В надстройке **Анализ Данных** выбрать инструмент анализа **Описательная статистика**.

2. В окне **Описательная статистика** в поле **Входной интервал:** ввести с помощью мышки диапазон ячеек **B1:B51** и нажать **Ввод**, выбрать **Группирование по столбцам**, поставить галочку в поле **Метки в первой строке** (в этом случае в заголовке результатов будет автоматически прописано «Масса клубней, г»), в поле **Выходной интервал** можно выбрать размещение на рабочем листе с исходными данными, Новый рабочий лист или Новая рабочая книга.

3. В случае размещения на листе 1 рядом с исходными данными в поле **Выходной интервал** необходимо указать интервал ячеек для размещения итоговых результатов описательной статистики, в нашем примере интервал **D16:E23**. Далее выбираем **Итоговую статистику**, в поле **Уровень надежности** указываем **95**, в полях **К-тый наименьший** и **К-тый наибольший** ставим **1** и нажимаем **ОК** (рис. 2.7)

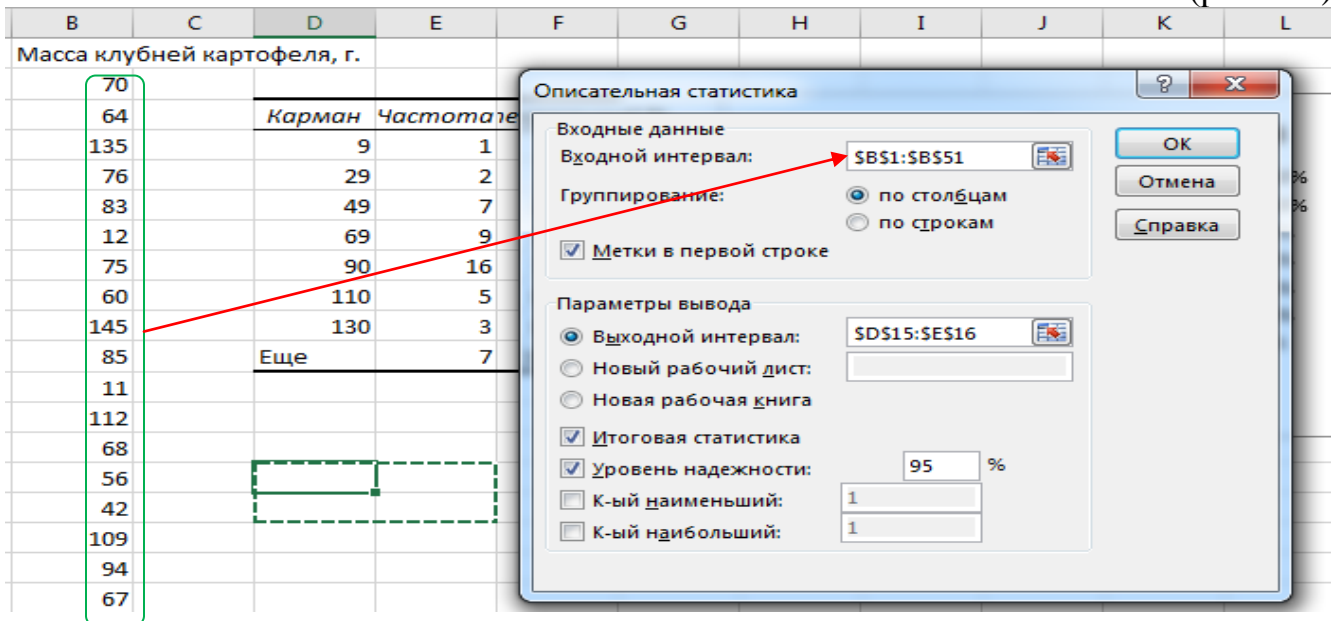


Рис. 2.7. Выбор параметров в окне *Описательная статистика*

4. После нажатия на клавишу **ОК** получаем результаты описательной статистики по массе 50 клубней картофеля (рис. 2.8).

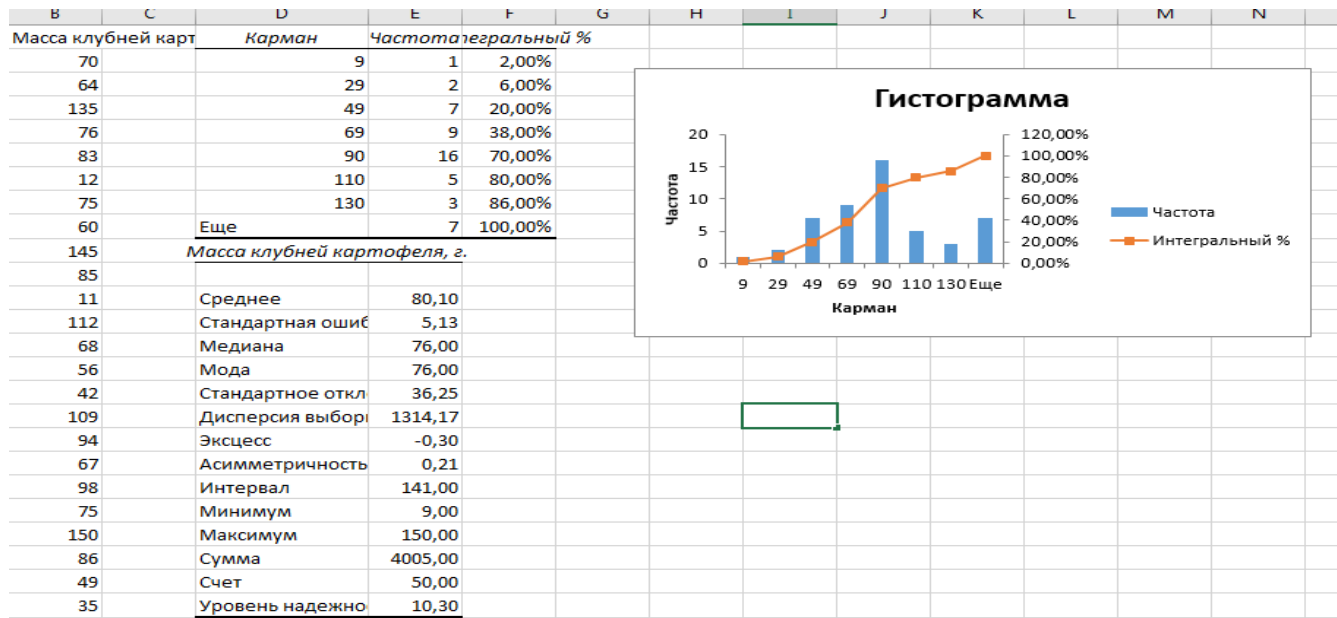


Рис.2.8. Исходные данные, таблица вариационного ряда, гистограмма и основные итоговые статистические показатели

Работа 3. Оценка двух вариантов при количественной изменчивости признаков

Теоретическая часть:

Вопрос о проверке гипотез – один из основных при проведении научных исследований в биотехнологии. Методы проверки гипотез являются надежной основой для принятия решения при некоторой неопределенности, обусловленной случайной вариацией изучаемых явлений и процессов. Они применяются всегда, когда необходимо решить вопрос о существенности разности между средними, о принадлежности варианта к данной совокупности, о значимости корреляционной зависимости между изучаемыми признаками и т.д.

Так, в частности, оценка разности средних изучаемых вариантов осуществляется в следующей последовательности:

- выдвигается нулевая гипотеза (H_0) – предположение об отсутствии реальных различий между средними значениями вариантов и альтернативная гипотеза (H_A) – предположение о наличии различий между вариантами;
- выбирается доверительная вероятность или уровень значимости для проверки нулевой гипотезы ($0,05$ – 5% или $0,01$ – 1%);
- выбирается критерий существенности (параметрический или непараметрический) для проверки гипотезы;
- рассчитываются статистические показатели для каждого варианта;
- в специальных таблицах находят теоретическое значение критерия для заданного уровня значимости;

- оценка разности средних (проверка H_0) может быть произведена по доверительным интервалам для генеральных средних, по фактическому значению критерия существенности или по наименьшей существенной разности (HCP_{05} , HCP_{01}).

После принятия или отбрасывания нулевой гипотезы необходимо сделать биологический вывод относительно изучаемых вариантов.

Общая постановка задачи:

1. Познакомиться с испытанием, как следующим после наблюдения видом научного исследования (пример из 2-х вариантов).
2. Освоить сущность и практическое применение статистических гипотез.
3. Подготовить данные к статистическому тесту и провести расчеты необходимых критериев для оценки существенности различий.
4. Построить в программе Excel гистограммы средних и доверительные интервалы генеральных средних (планку погрешностей) или Box-plot.
5. Сделать правильное статистическое и биологическое заключение по эффективности изученных вариантов.

Контрольные вопросы:

1. Различия между зависимыми и независимыми выборками на примерах.	3. Нулевая гипотеза и статистические методы ее проверки.
2. Оценка существенности разности средних зависимых выборок.	4. Параметрические и непараметрические критерии для проверки гипотез.

А. Независимые выборки

Пример 1. Изучается два варианта режима клонального микроразмножения оздоровленных растений картофеля *in vitro*: 1 – среда Мурасиге-Скуга (MS). 2 – среда Гамборга (B5). Изучаемый признак высота стебля на 21 день, см.

Таблица 6

Среда Мурасиге-Скуга (вариант 1)					Среда Гамборга (вариант 2)				
Номера колонок					Номера колонок				
1	2	3	4	5	6	7	8	9	10
15,6	12,4	10,2	8,0	9,4	17,8	13,5	12,5	13,0	12,4
14,4	10,6	11,0	9,8	8,8	16,6	14,2	13,6	14,4	13,7
13,9	9,3	8,1	9,1	7,4	17,0	15,5	10,4	12,8	14,3
16,0	13,4	9,7	7,3	10,4	15,6	14,5	11,8	12,0	14,5
14,9	12,6	7,6	10,0	6,5	16,5	13,0	13,9	11,5	12,5
13,3	10,5	8,4	7,5	9,9	17,0	14,3	14,8	12,8	13,7
14,4	10,5	9,0	11,1	9,0	17,1	13,6	11,3	13,5	12,9
10,4	9,0	9,7	10,0	9,6	16,4	12,9	12,5	14,4	14,0
11,1	8,8	9,5	10,3	7,4	19,3	15,4	10,4	15,9	14,5
15,0	9,4	–	8,4	9,4	16,7	13,9	–	11,8	13,4

Пример 2. Гибель свекловичного долгоносика при обработке: биотехнологическим препаратом на основе *Bacillus thuringiensis* – энтобактерин-3 (вариант 1) и фозалоном (вариант 2) составила (%):

Таблица 7

Энтобактерин-3 (вариант 1)					Фозалон (вариант 2)				
Номера колонок					Номера колонок				
1	2	3	4	5	6	7	8	9	10
95	78	89	96	87	67	78	89	56	67
67	78	90	89	94	67	58	79	76	57
70	89	95	95	86	68	59	73	69	75
69	68	85	82	76	75	64	82	48	82
77	85	82	70	81	59	66	75	46	57
79	87	69	98	87	65	45	67	70	75
70	87	78	79	69	56	60	59	64	45
83	86	87	80	79	67	64	54	50	60
92	94	95	76	82	65	67	46	65	67
90	78	80	–	67	53	57	70	46	–

Таблица 8

Шифр задания к работе № 3

№ задания	колонки	№ задания	колонки	№ задания	колонки	№ задания	колонки
1	1 6	8	2 8	15	4 7	22	5 9
2	1 7	9	2 9	16	4 8	23	5 10
3	1 8	10	2 10	17	4 9	24	3 10
4	1 9	11	3 6	18	4 10	25	4 6
5	1 10	12	3 7	19	5 6	26	2 10
6	2 6	13	3 8	20	5 7	27	3 10
7	2 7	14	3 9	21	5 8	28	5 10

Статистическая обработка:

Таблица 9

Рабочая таблица с формулами для расчетов

Статистические показатели	Вариант 1	Вариант 2
$\bar{x} = \frac{\sum X}{n} =$		
$C = \frac{(\sum X)^2}{n} =$		
$CK = \sum (X - \bar{x})^2 = \sum X^2 - C =$		
$S^2 = \frac{\sum (X - \bar{x})^2}{n-1} =$		
$S = \sqrt{S^2} =$		
$s_x = \frac{S}{\sqrt{n}} =$		
$V = \frac{S}{\bar{x}} \cdot 100 =$		

Оценка существенности разности между средними по вариантам

а) интервальным методом

$$\bar{x}_1 \pm t_{05} \cdot s_{\bar{x}_1} =$$

$$t_{05} =$$

$$\bar{x}_2 \pm t_{05} \cdot s_{\bar{x}_2} =$$

$$\text{при } df = n-1$$

б) по критерию существенности

$$t_{\delta} = \frac{d}{S_d} = \frac{|\bar{x}_2 - \bar{x}_1|}{\sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2}} =$$

$$t_{05} = \quad t_{01} =$$

$$\text{при } df = n_1 + n_2 - 2 =$$

в) по наименьшей существенной разности (НСР)

$$НСР_{05} = t_{05} \cdot S_d =$$

$$НСР_{01} = t_{01} \cdot S_d =$$

Заключение:

Оценка двух вариантов при количественной изменчивости признаков для независимых выборок в программе Excel

Эту работу можно выполнить 2 инструментами **Пакета данных**: двухвыборочным t-тестом с одинаковыми дисперсиями или с помощью описательной статистики.

1. Двухвыборочный t-тест с одинаковыми дисперсиями

Пример 1. Содержание белка (%) в зерне при испытании двух сортов озимой пшеницы Престиж и Немчиновская-57

Престиж	17,4	18,6	20,3	18,4	19,6	19,5	19,5	19,0	18,8	19,4
Немчиновская -57	18,0	16,4	16,5	17,0	17,5	16,8	18,5	17,2	15,9	16,8

1. В активный лист программы Excel в ячейках столбцов **A** и **B** ввести тексты «Престиж», «Немчиновская-57» и данные по содержанию белка (рис. 3.1).

2. В надстройке **Анализ данных** выбрать инструмент анализа **Двухвыборочный t-тест с одинаковыми дисперсиями** (рис.3.1).

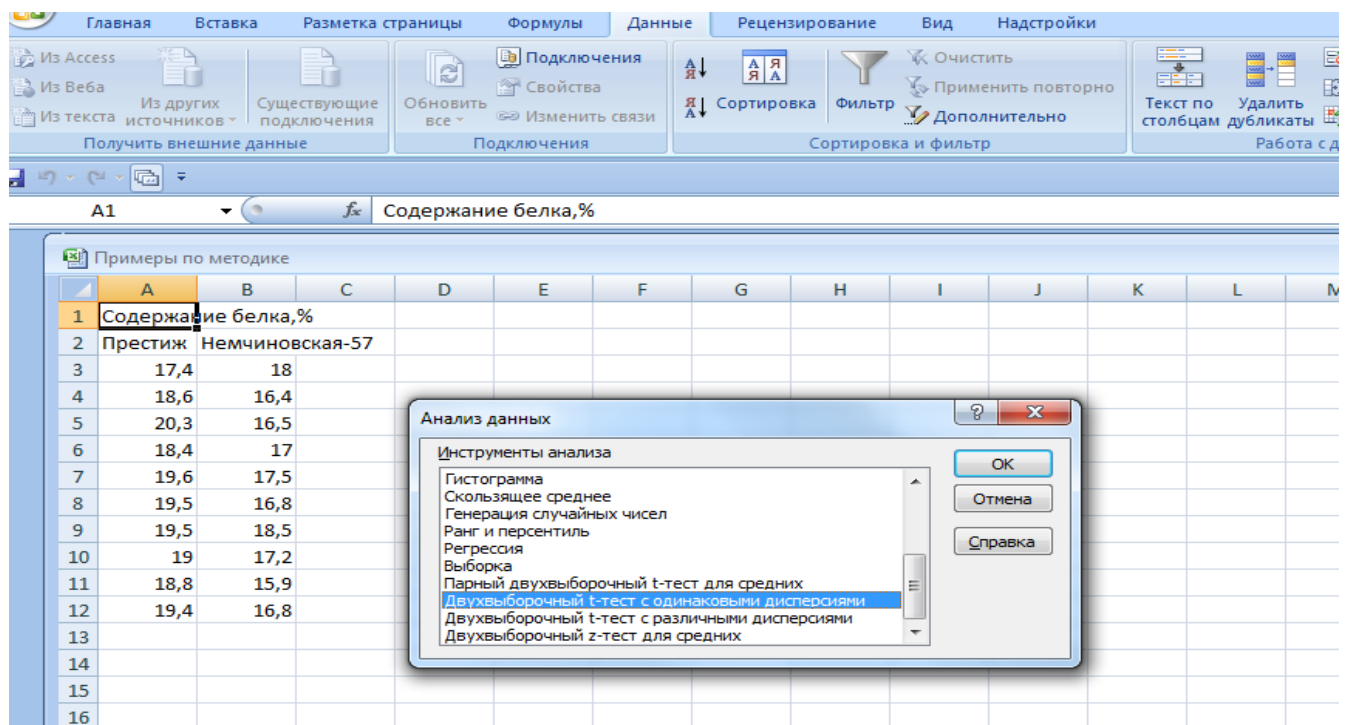


Рис. 3.1. Таблица с исходными данными по содержанию белка и выбор инструмента анализа

2. В окне **Двухвыборочный t-тест с одинаковыми дисперсиями** в поле **Интервал переменной 1:** ввести с помощью левой кнопки мыши диапазон ячеек **A2:A12** и нажать **Ввод**, в поле **Интервал переменной 2:** ввести с помощью левой кнопки мыши диапазон ячеек **B2:B12** и нажать **Ввод**. В поле **Гипотетическая разность** указать **0**, поставить галочку в поле **Метки**. В поле **Альфа** указать **0,05**, выбрать место размещения в выходном интервале. После заполнения всех полей нажимаем **ОК** (рис. 3.2.)

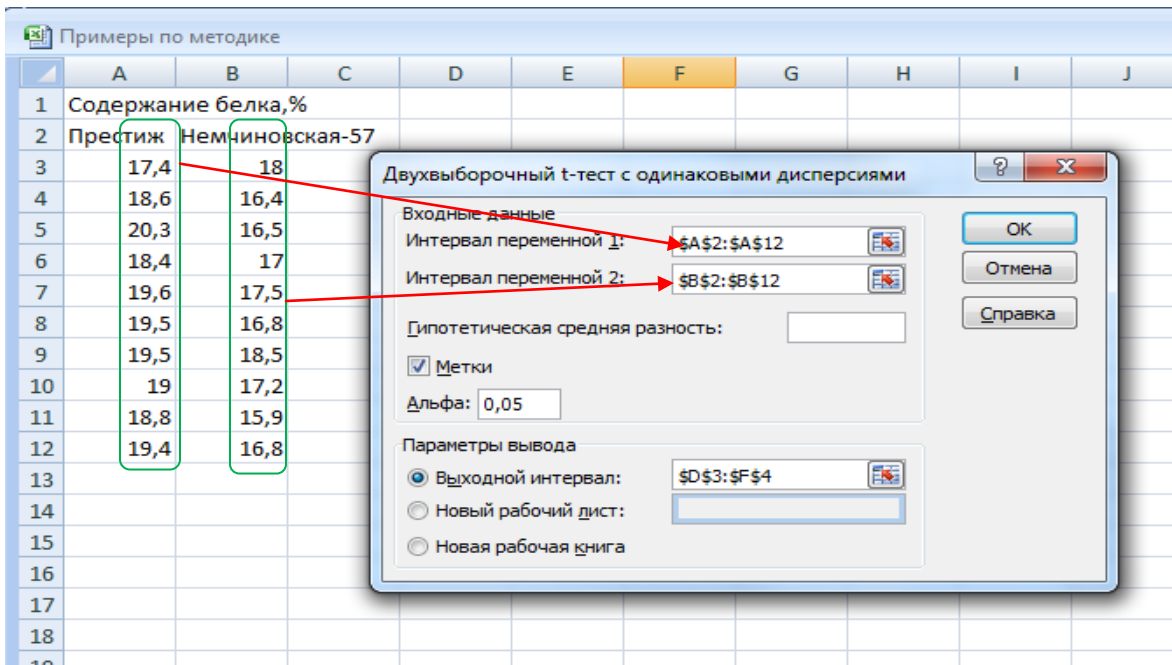


Рис. 3.2. Диалоговое окно *Двухвыборочный t-тест с одинаковыми дисперсиями*

4. После нажатия на клавишу **ОК** получаем итоговую таблицу (рис. 3.3), в которой самыми важными показателями являются фактическое значение критерия Стьюдента ($t_{\text{фак}} - t$ -статистика) и табличное – t_{05} (критическое). Так как фактическое значение критерия Стьюдента (t -статистика = 5,64) в нашем примере больше табличного (t -критическое двухстороннее = 2,10 при числе степеней свободы $df=10+10-2 = 18$), H_0 отвергается. С 95% вероятностью можно предполагать, что сорт озимой пшеницы Престиж содержит существенно больше белка, чем сорт Немчиновская-57.

Примеры по методике						
	A	B	C	D	E	F
1	Содержание белка, %					
2	Престиж	Немчиновская-57				
3	17,4	18		Двухвыборочный t-тест с одинаковыми дисперсиями		
4	18,6	16,4				
5	20,3	16,5			Престиж	Немчиновская-57
6	18,4	17		Среднее	19,05	17,06
7	19,6	17,5		Дисперсия	0,645	0,600444444
8	19,5	16,8		Наблюдения	10	10
9	19,5	18,5		Объединенная дисперсия	0,62272222	
10	19	17,2		Гипотетическая разность средних	0	
11	18,8	15,9		df	18	
12	19,4	16,8		t-статистика	5,6388546	
13				P(T<=t) одностороннее	1,1904E-05	
14				t критическое одностороннее	1,73406359	
15				P(T<=t) двухстороннее	2,3809E-05	
16				t критическое двухстороннее	2,10092204	
17						
18						
19						

Рис.3.3. Итоговая таблица двухвыборочного t-теста

2. Описательная статистика

1. В активный лист программы Excel в ячейках столбцов **A** и **B** ввести текст «Престиж», «Немчиновская-57» и данные по содержанию белка (рис.3.4).
2. В надстройке **Анализ данных** выбрать инструмент анализа **Описательная статистика** (рис.1.11)
3. Выбрать параметры окна **Описательная статистика** (рис.3.4)

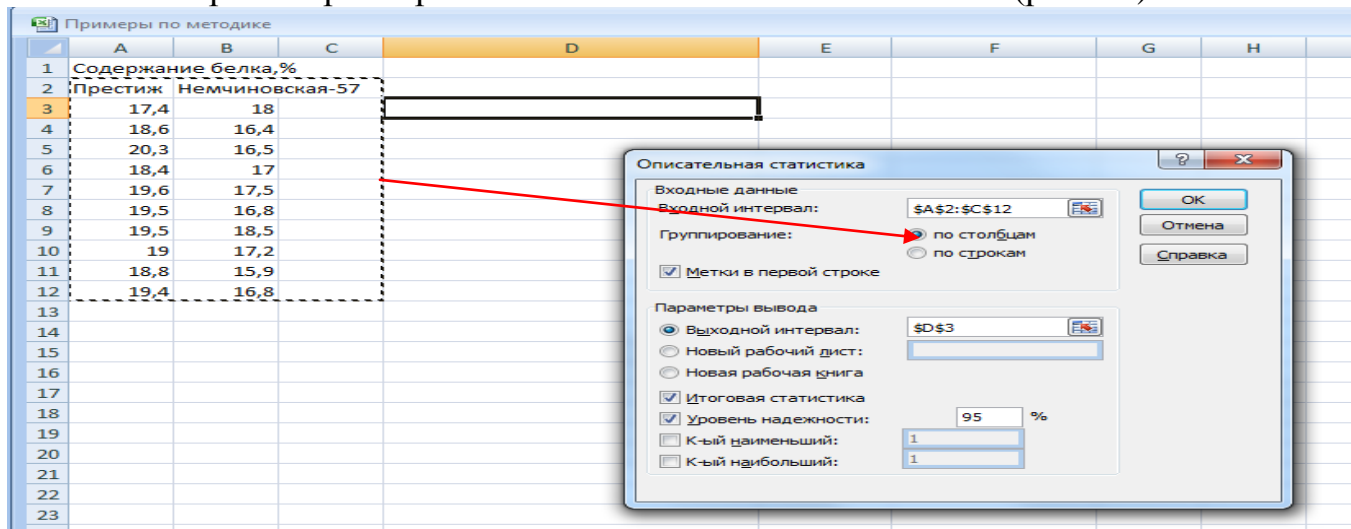


Рис. 3.4. Выбор параметров описательной статистики

3. После нажатия на клавишу **Ок** получаем результаты основных статистических показателей по каждой выборке (сорт) (рис. 3.5).

	A	B	C	D	E	F	G	H	I
1	Содержание белка,%								
2	Престиж	Немчиновская-57							
3	17,4	18		<i>Престиж</i>		<i>Немчиновская-57</i>			
4	18,6	16,4							
5	20,3	16,5		Среднее	19,05	Среднее	17,06		
6	18,4	17		Стандартная ошибка	0,25	Стандартная ошибка	0,25		
7	19,6	17,5		Медиана	19,20	Медиана	16,90		
8	19,5	16,8		Мода	19,50	Мода	16,80		
9	19,5	18,5		Стандартное отклонение	0,80	Стандартное отклонение	0,77		
10	19	17,2		Дисперсия выборки	0,64	Дисперсия выборки	0,60		
11	18,8	15,9		Эксцесс	1,05	Эксцесс	0,04		
12	19,4	16,8		Асимметричность	-0,68	Асимметричность	0,55		
13				Интервал	2,90	Интервал	2,60		
14				Минимум	17,40	Минимум	15,90		
15				Максимум	20,30	Максимум	18,50		
16				Сумма	190,50	Сумма	170,60		
17				Счет	10,00	Счет	10,00		
18				Уровень надежности(95,0%)	0,57	Уровень надежности	0,55		
19									

Рис. 3.5. Итоги описательной статистики

Для оценки существенности разности средних двух сортов озимой пшеницы по содержанию белка наибольший интерес представляют следующие данные (они выделены красным шрифтом): среднее содержание белка у сорта Престиж, $\bar{x}_1 = 19,05\%$, у сорта Немчиновская-57, $\bar{x}_2 = 17,06\%$; ошибки выборочной средней

(стандартные ошибки), соответственно, $S_{x_1} = 0,25$ и $S_{x_2} = 0,25\%$, а также предельные ошибки выборочной средней (в **Пакете данных** – уровень надежности), соответственно, 0,57 и 0,55. Оценку существенности разности средних проведем по доверительным интервалам для генеральных средних:

$$\bar{x}_1 \pm t_{0,5} \cdot s_{\bar{x}_1} = 19,05 \pm 0,57 = 18,48 \div 19,62\%$$

$$\bar{x}_2 \pm t_{0,5} \cdot s_{\bar{x}_2} = 17,06 \pm 0,55 = 16,51 \div 17,61\%$$

Вывод: так как доверительные интервалы для генеральных средних не перекрываются, с вероятностью 95% можно предполагать, что сорт озимой пшеницы **Престиж** содержит существенно больше белка, чем сорт **Немчиновская-57**.

Графическое представление средних с ошибками

1. Построение графиков с погрешностями

1. Выбрать с помощью левой кнопки мышки диапазон **A14:B15**, включающий наименование сортов и их средние значения, затем в строке меню нажать на вкладку **Вставить**, далее – один из видов **Диаграммы**, например, **Гистограмма**. После нажатия на клавишу **Ок** появляются столбчатые гистограммы среднего количества белка у испытуемых сортов озимой пшеницы.

2. Далее в строке меню **Работа с диаграммами** необходимо выбрать подменю **Конструктор** и **Макет**. Инструменты данных подменю дают возможность провести форматирование гистограммы (название диаграммы, подписи данных, масштабирование осей и т.д. (Рис. 3.6).

3. Для выделения на гистограмме погрешностей в программах Excel – 2013, 2016 в окне **ЭЛЕМЕНТЫ ДИАГРАММЫ** галочкой выделяем **Предел погрешностей**, в программе Excel 2007 – выбираем **Планки погрешностей** (Рис. 3.6).

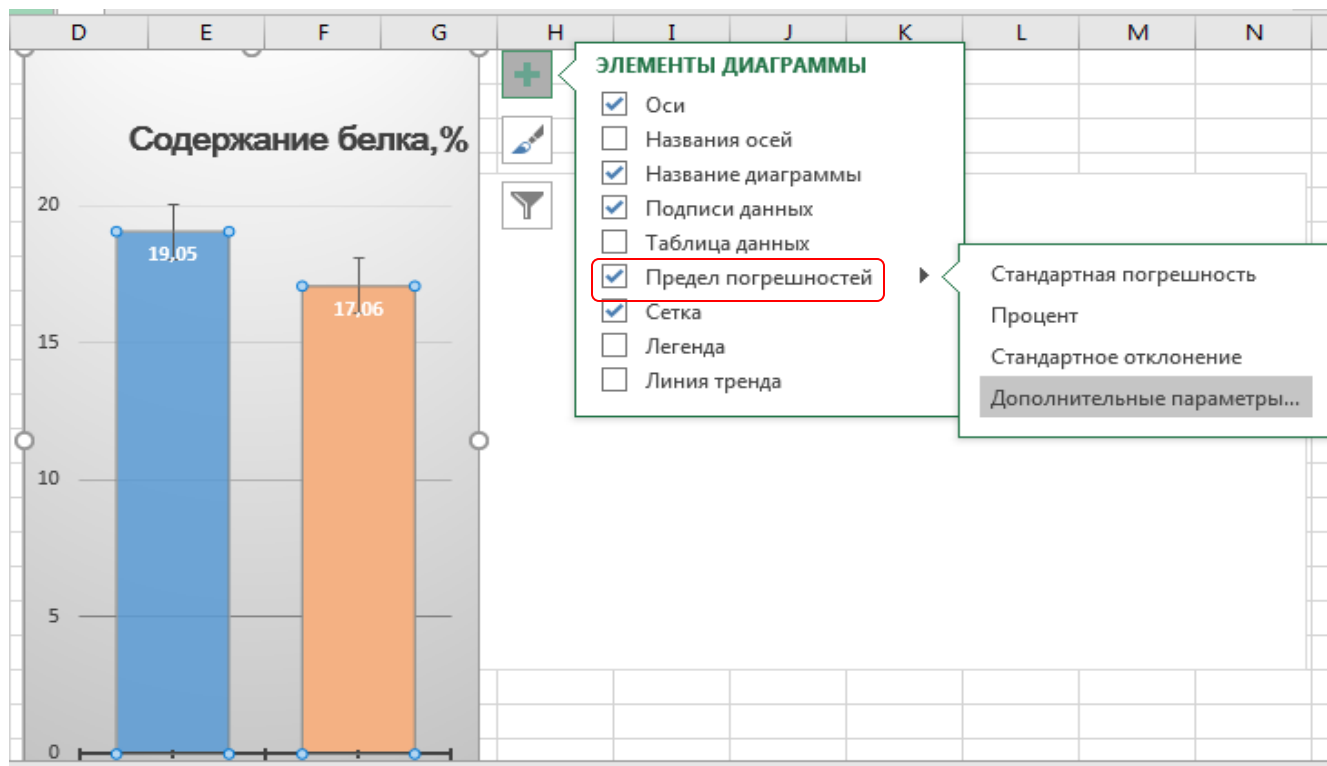


Рис. 3.6. Выбор предела погрешностей в Excel – 2016

4. На графике над каждым столбиком появляются вертикальные линии, при этом по умолчанию высота этих линий равна стандартной погрешности, причем они одинаковые по размерам для всех сравниваемых вариантов.

5. После нажатия мышкой на **Предел погрешностей**, далее **Дополнительные параметры** появляется выпадающее окно **Формат предела погрешностей** (рис.3.7), в котором можно выбрать параметры предела погрешностей: цвет, тип и величина погрешностей.

6. В программе Excel можно установить следующие типы полос (пределов) погрешностей (рис. 3.7):

- **Фиксированное значение.** Полосы погрешностей откладываются от каждой точки данных на заданную пользователем фиксированную величину. Все полосы погрешностей имеют одинаковую высоту.

- **Относительное значение.** Полосы погрешностей откладываются от каждой точки данных на величину, определяемую в процентах от значения точки. Например, если задать в поле ввода значение 5%, а значение точки равно 60, то полоса погрешности будет выведена от 57 до 63.

- **Стандартное отклонение.** Полосы погрешностей откладываются на величину стандартного отклонения – S . Все полосы погрешностей имеют одинаковую высоту и откладываются вверх и вниз на величину S не от конкретной точки, а от среднего всех точек.

- **Стандартная погрешность.** Размер полос погрешностей одинаковый для всех точек, и откладываются полосы погрешностей от среднего значения каждой выборки, а не от среднего значения по всем точкам.

- Пользовательская. Полосы погрешностей в зависимости от поставленных задач определяет пользователь и значения этих полос хранятся в заданном диапазоне, который может содержать формулы.

7. Так как предельные ошибки выборочных средних по вариантам могут иметь разные значения, рекомендуется выбирать не фиксированные значения, а указывать фактические данные. По сути дела, с помощью планок погрешностей нам мы можем показать на графике величину доверительных интервалов для генеральных средних двух сортов пшеницы:

$$\bar{x}_1 \pm t_{05} \cdot s_{\bar{x}_1} = 19,05 \pm 0,57 = 18,48 \div 19,62\% \quad \bar{x}_2 \pm t_{05} \cdot s_{\bar{x}_2} = 17,06 \pm 0,55 = 16,51 \div 17,61\%$$

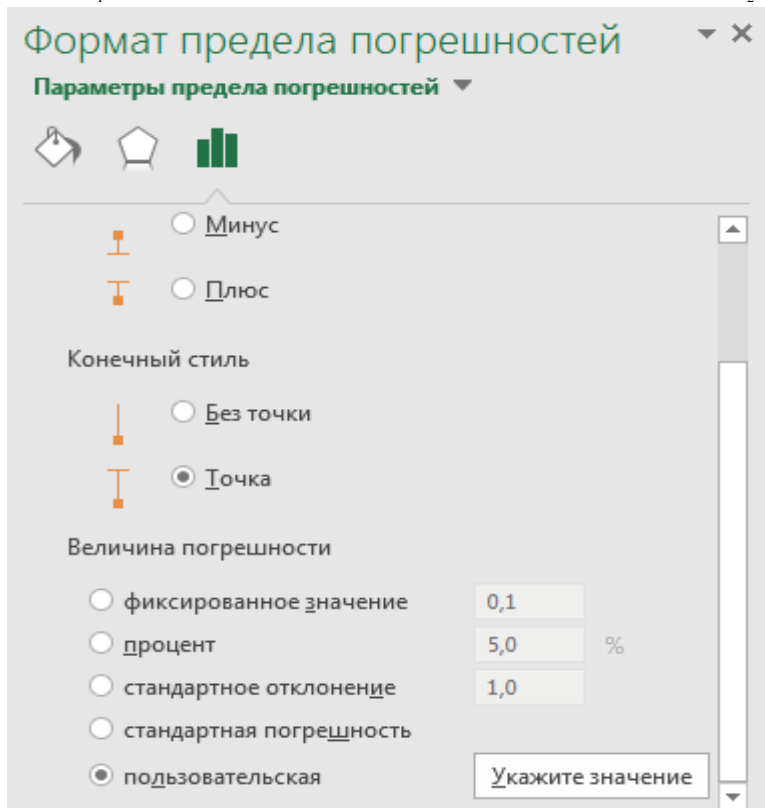
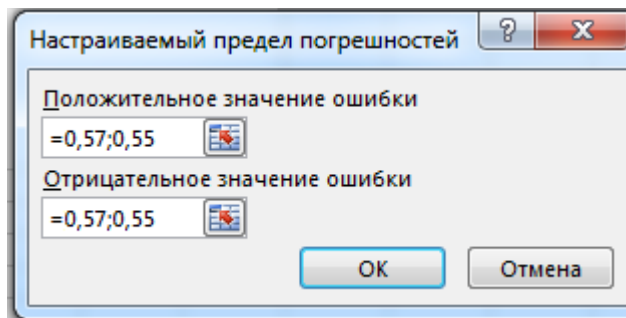


Рис. 3.7. Формат предела погрешностей

Для этого выберем формат **Пользовательская** (рис. 3.7) и в выпадающей панели **Настраиваемый предел погрешностей** для каждого ряда указываем величину предельной ошибки следующим образом: в окошко «положительное значение ошибки» через точку с запятой: 0,55; 0,57.



В окошко «отрицательное значение ошибки» вводится тоже самое или указываем ссылки на ячейки, где предельные ошибки рассчитаны для каждого варианта (рис.3.8).

Рис. 3.8. Панель для указания значений предельной ошибки

8. После завершения необходимых операций с **Планками погрешностей** получаем гистограмму среднего содержания белка у сортов озимой пшеницы

Престиж и Немчиновская – 57 с вертикальными линиями, которые отражают по сути дела графически величину доверительных интервалов для генеральных средних (рис. 3.9).

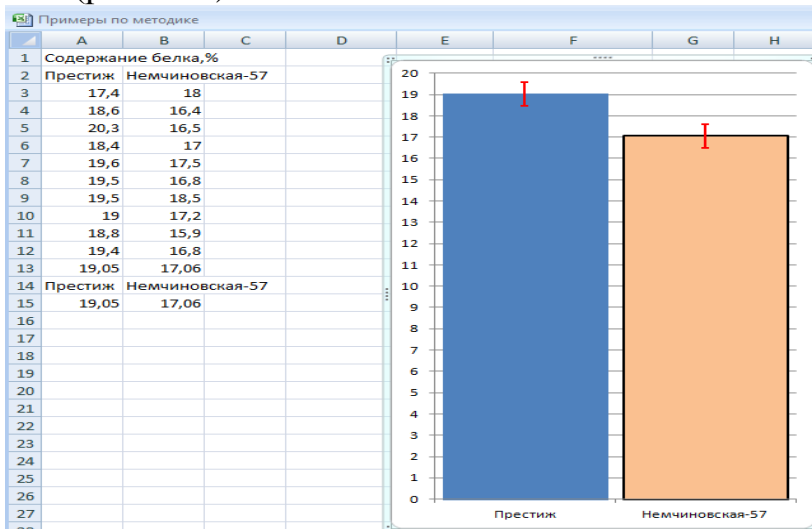


Рис. 3.9. Гистограмма с пользовательскими пределами (полосами) погрешностей

9. Так как верхняя граница предела погрешности (доверительного интервала) у сорта Немчиновская – 57 не пересекается с нижним пределом погрешности (доверительного интервала) у сорта Престиж, это свидетельствует о том, что между этими сортами с 95% вероятностью есть существенные различия по содержанию белка (рис. 3.9).

10. Гистограмму с пределами погрешностей можно преобразовать в любой тип диаграммы. Для этого необходимо подвести курсор мышки к гистограмме и, нажав на правую клавишу мышки выбрать различные виды диаграмм, в нашем примере произведен выбор точечной диаграммы (ромбики указывают величину средних значений, а вертикальные линия обозначают границы доверительных интервалов для генеральных средних (рис. 3.10).

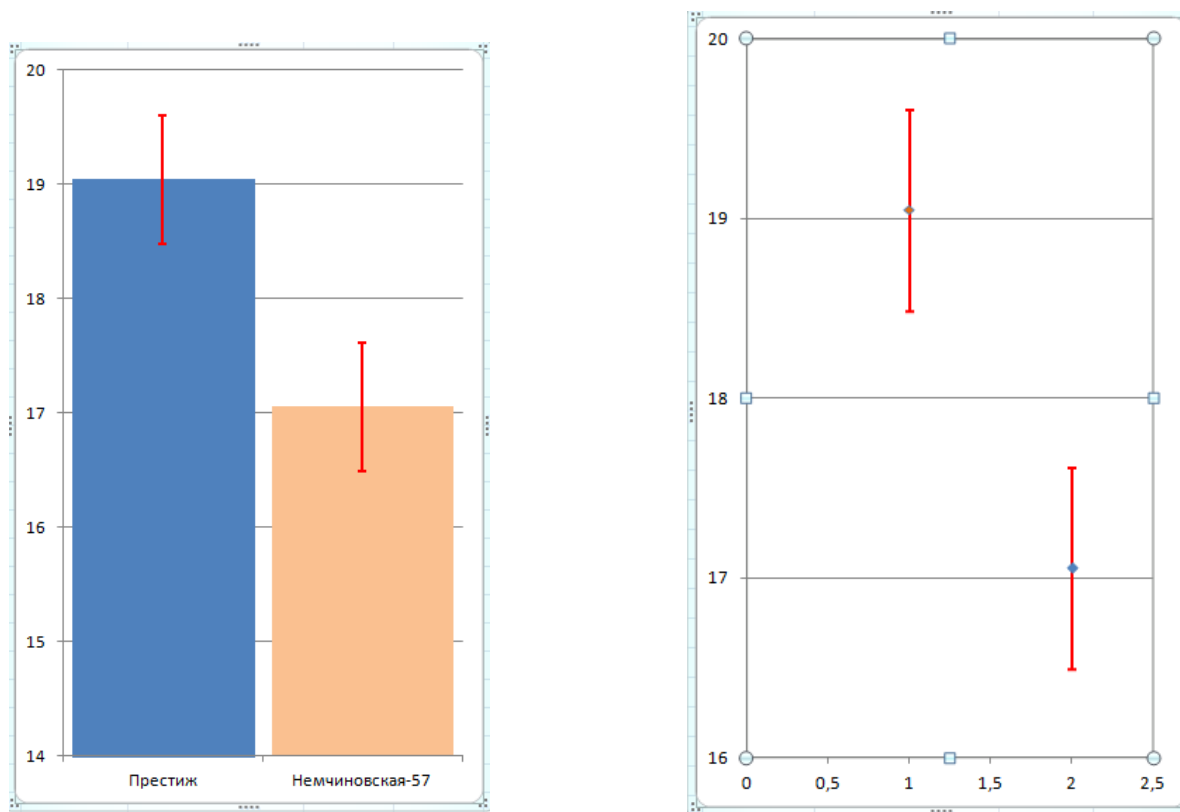


Рис. 3.10. Гистограмма и точечный график средних и доверительных интервалов

2. Построение графика «Ящик с усами – Boxplot»

В программах Excel 2016 и выше возможно построение графиков **Box-Plot** «ящик с усами» или «коробчатая диаграмма» (box-and-whiskers plot или boxplot), в которых графически отображаются средняя арифметическая, медиана и доверительные интервалы. Такие диаграммы очень удобны при графическом сравнении 2-х или большего числа вариантов между собой.

Для построения указанной диаграммы обратимся к нашему примеру по содержанию белка у сортов Престиж и Немчиновская-57.

1. Выберем с помощью левой кнопки мышки диапазон **A14:B15**, включающий наименование сортов и данные по содержанию белка, затем в строке меню нажимаем на вкладку **Вставить**, далее – в списке рекомендуемых диаграмм отметим **Ящик с усами**. После нажатия на клавишу **Ok** появляются стандартные формы 2-х коробчатых диаграмм, которые необходимо настроить (рис. 3.11).

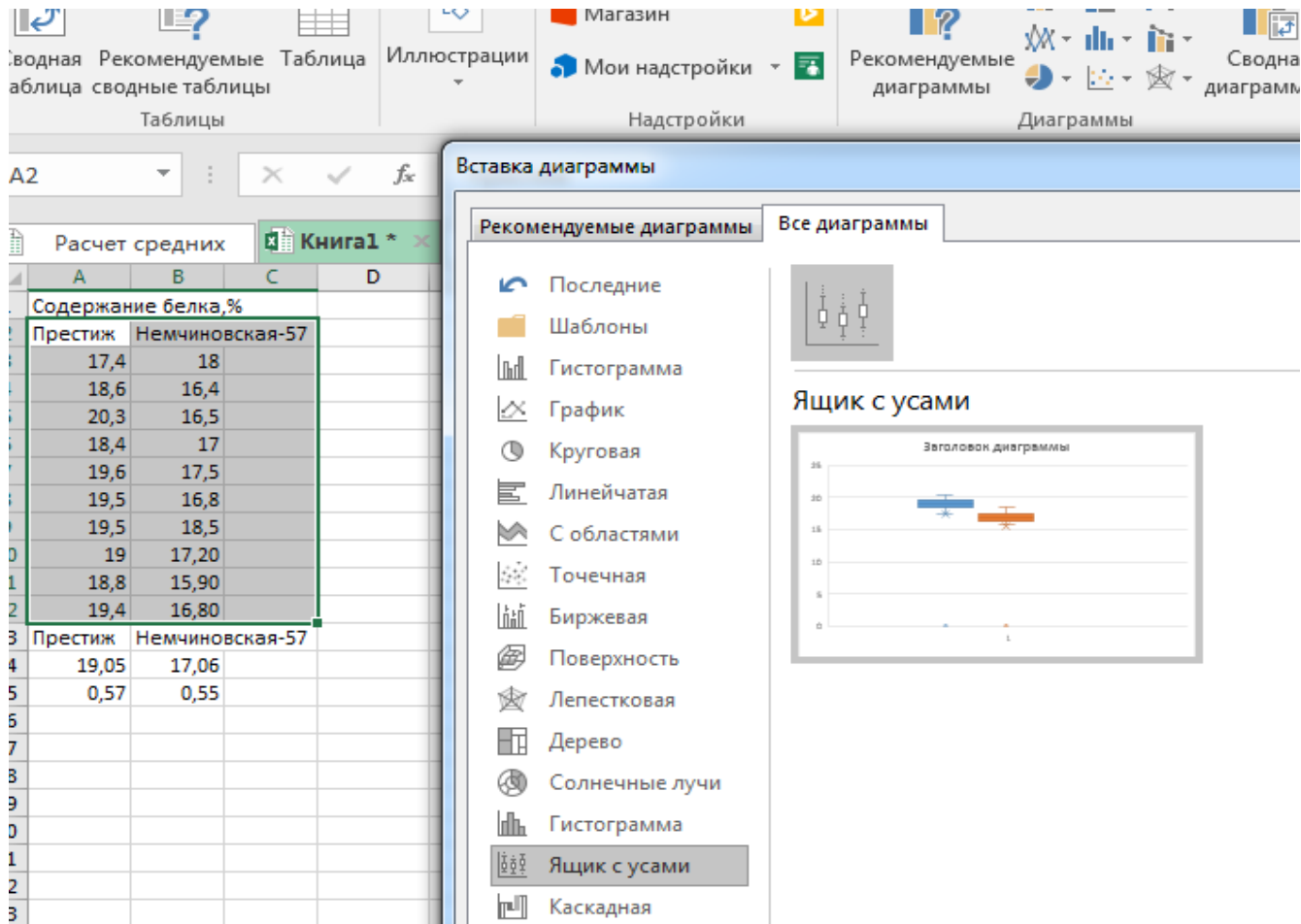


Рис.3.11. Выбор диаграммы «Ящик с усами» в программе Excel – 2016

2. Общий вид диаграммы настраивается стандартно. Можно менять цвет, добавлять подписи и т.д. Для этого есть две контекстные вкладки **Конструктор** и **Формат** или можно воспользоваться настройками, предназначенными специально для этой диаграммы. Для этого выбираем какой-либо ряд и нажимаем на **Ctrl+1** или левой кнопкой мышки подводим курсор на любой «ящик», справа появляется панель настроек **Формат ряда данных** (рис.3.12).

3. С помощью настроек можно форматировать следующие параметры:

Боковой зазор – регулирует ширину ящиков и расстояние между ними.

Показывать внутренние точки. Если поставить галочку, то на оси, где расположены «усы», точками будут показаны все значения по данному ряду.

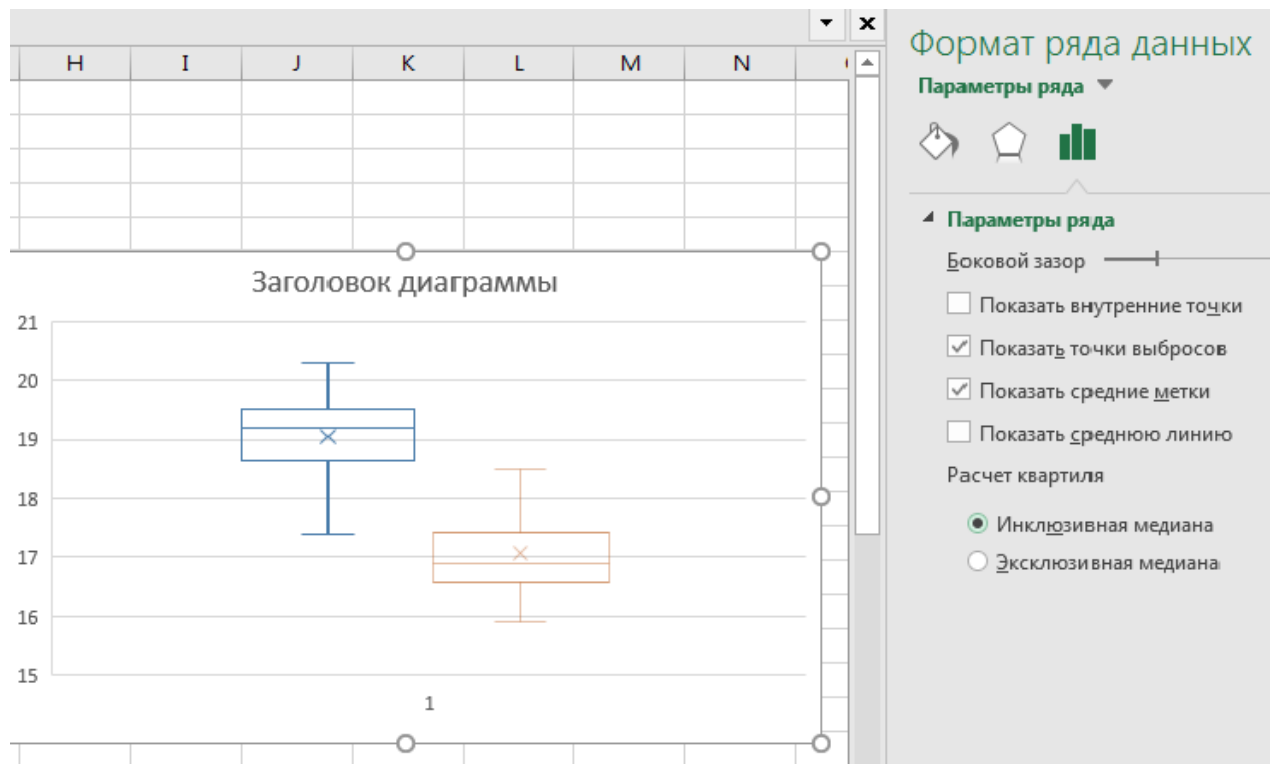


Рис.3.12. Настройка диаграммы «Ящик с усами»

Показать средние метки – среднее арифметическое (крестики).

Показать среднюю линию – только для различных категорий.

Показывать точки выбросов – отражать экстремальные значения.

Инклюзивная медиана или **экслюзивная медиана**. Инклюзивная медиана включает квартильные значения, а экслюзивная медиана не включает.

4. Наждем на знак + в правом углу рабочей области диаграммы, появляется окно **ЭЛЕМЕНТЫ ДИАГРАММЫ**, в котором галочками отметим нужные для нас параметры диаграммы. После настроек параметров «Ящика с усами» получаем следующий график (рис. 3.13).

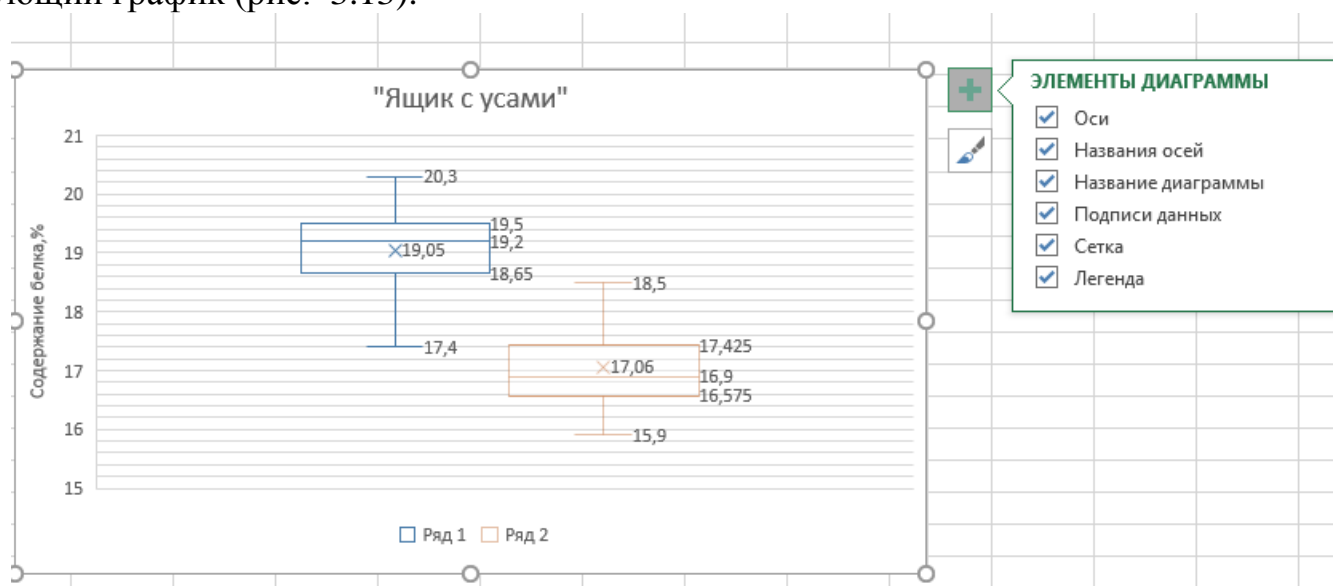


Рис. 3.13. «Ящики с усами» содержания белка у сортов Престиж (Ряд 1) и Немчиновская-57 (Ряд2)

5. На диаграммах показано:

Крестик посередине – это среднее арифметическое или среднее выборочное по каждому сорту: 19,05 и 17,06.

Линия чуть выше или ниже крестика – медиана для сорта Престиж – 19,2 и 16,9 – для сорта Немчиновская-57.

Нижняя и верхняя грань прямоугольника (типа ящика) соответствует первому и третьему квартилю (значениям, отделяющим $\frac{1}{4}$ и $\frac{3}{4}$ выборки). Расстояние между 1-м и 3-м квартилем – это межквартильный размах (или расстояние).

Горизонтальные черточки на конце «усов» – максимальное и минимальное значение (без учета выбросов).

По представленным диаграммам в виде «ящичков с усами» можно проверить нулевую гипотезу об отсутствии различий между выборками (вариантами). При этом тестируемые данные могут подчиняться нормальному распределению или же отличаться от нормального распределения. Сравнение проводится по значениям квартилей: если нижняя грань прямоугольника, соответствующая первому квартилю одного варианта, заходит за верхнюю грань другого прямоугольника, соответствующему третьему квартилю другого варианта, нулевая гипотеза принимается, и различия не существенны.

В нашем примере, так как нижний квартиль сорта Престиж равен 18,65, а верхний квартиль сорта Немчиновская – 57 равен 17,425 (ящички не прикасаются между собой), значит можно предполагать, что различия по содержанию белка между указанными сортами значимы на 0,05 уровне значимости.

В. Зависимые (сопряженные) выборки

Пример 1. По результатам исследования влияния освещенности на активность пероксидазы установлено следующее количество фермента в листьях верхнего и нижнего ярусов (мл 0,01 н. раствора на 1 г сырого вещества, n = 10):

Нижний	5,7	6,0	7,3	6,9	6,4	8,1	5,4	6,5	6,8	7,2
Верхний	7,8	7,5	8,0	8,5	8,9	8,3	7,4	6,9	8,5	8,2

Пример 2. В колбы Эрленмейера вместимостью 250 мл, содержащие 100 мл стерильной морской воды, вносили сырую стерильную нефть (1% по объему) и 1,0 мл суспензии штамма *Phyllobacterium myrsinacearum* ($2,0 \times 10^6$ кл./мл). Контролем служили колбы со стерильной морской водой с добавлением стерильной сырой нефти (1 % по объему) без добавления штамма.

Колбы инкубировали на одном и том же перемешивающем устройстве.

Угледородокисляющую активность штамма оценивали по изменению суммарного количества углеводов на приборе Флюорат-02-3 М

Варианты	Убыль углеводов, % на 10 день									
	Морская вода	36	28	34	25	21	36	25	23	43
Морская вода с добавлением <i>Ph. myrsinacearum</i>	55	53	65	60	48	55	56	24	40	60

Статистическая обработка:

Таблица 10

Пары	Вариант 1	Вариант 2	Разность $d = X_2 - X_1$	Квадрат разности d^2
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
Суммы			$\Sigma d =$	$\Sigma d^2 =$
			$\bar{d} = \frac{\Sigma d}{n} =$	—

$$1. S_{\bar{d}} = \sqrt{\frac{\Sigma d^2 - \frac{(\Sigma d)^2}{n}}{n \cdot (n-1)}} =$$

$$2. t_{\phi} = \frac{\bar{d}}{S_{\bar{d}}} = \quad \text{при } df = n - 1 = \quad t_{05} = \quad t_{01} =$$

$$3. HCP_{05} = t_{05} \cdot S_{\bar{d}} = \quad HCP_{01} = t_{01} \cdot S_{\bar{d}} =$$

Заключение:

Оценка средней разности при количественной изменчивости признаков для зависимых выборок в программе Excel

Парный двухвыборочный t-тест для средних.

Пример 2. При изучении 2-х способов хранения яблок в полиэтиленовых пакетах в одних и тех же камерах холодильника процент сохранившихся плодов составил (% , $n = 10$)

Без газовой среды	56	68	74	75	80	56	63	66	75	64
Газовая среда	85	73	75	95	78	85	76	74	83	60

1. В активный лист программы Excel введем исходные данные вышеприведенного примера, расположив таблицу по столбцам (рис. 3.14.).

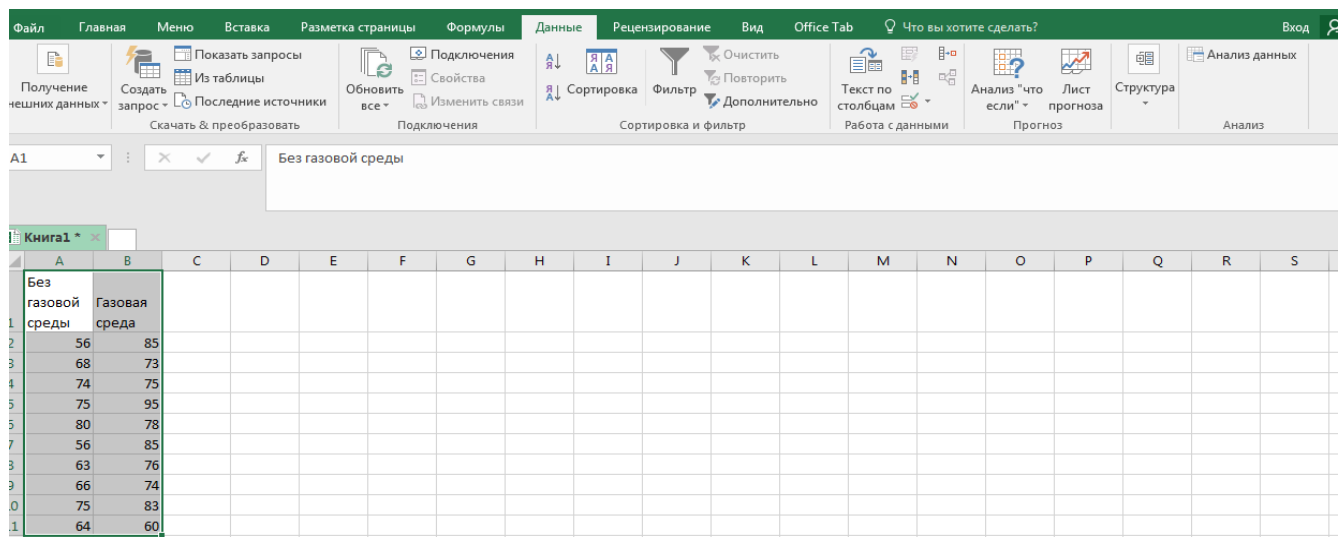


Рис. 3.14. Исходные данные

2. В надстройке Анализ данных выбрать инструмент анализа Парный двухвыборочный t-тест для средних (рис.3.15)

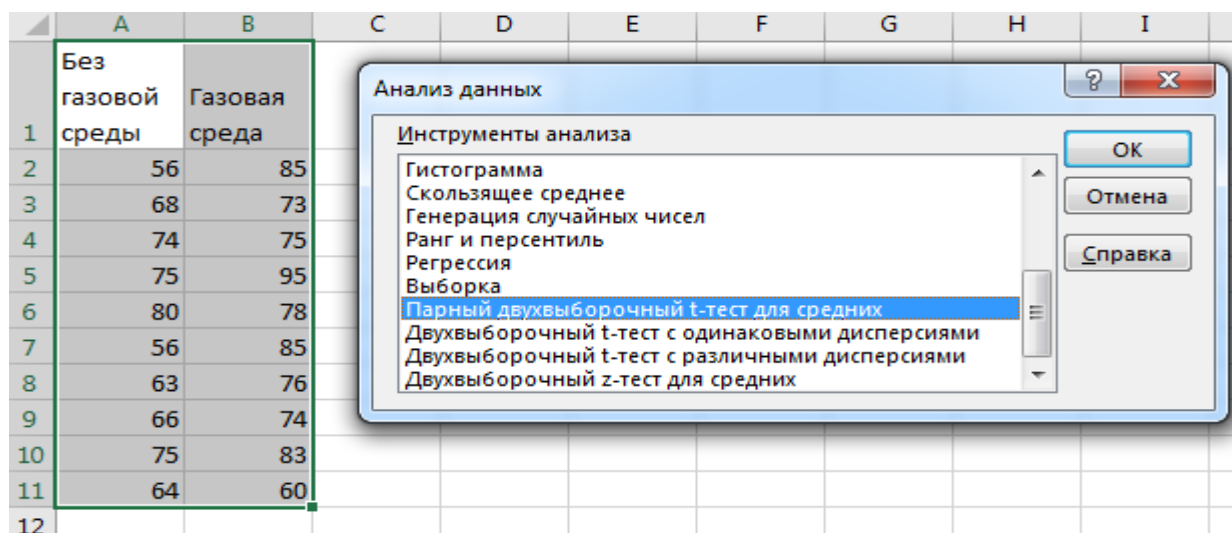


Рис. 3.15. Подменю Анализ данных

В появившемся окне **Парный двухвыборочный t-тест для средних** в поле **Интервал переменной 1:** ввести с помощью мышки диапазон ячеек **A1:A11** и нажать **Ввод**, в поле **Интервал переменной 2:** ввести с помощью мышки диапазон ячеек **B1:B11** и нажать **Ввод**. В поле **Гипотетическая разность** указать **0**, поставить галочку в поле **Метки**. В поле **Альфа** указать **0,05**, выбрать место размещения в выходном интервале. После заполнения всех полей нажимаем **ОК** (рис. 3.16.)

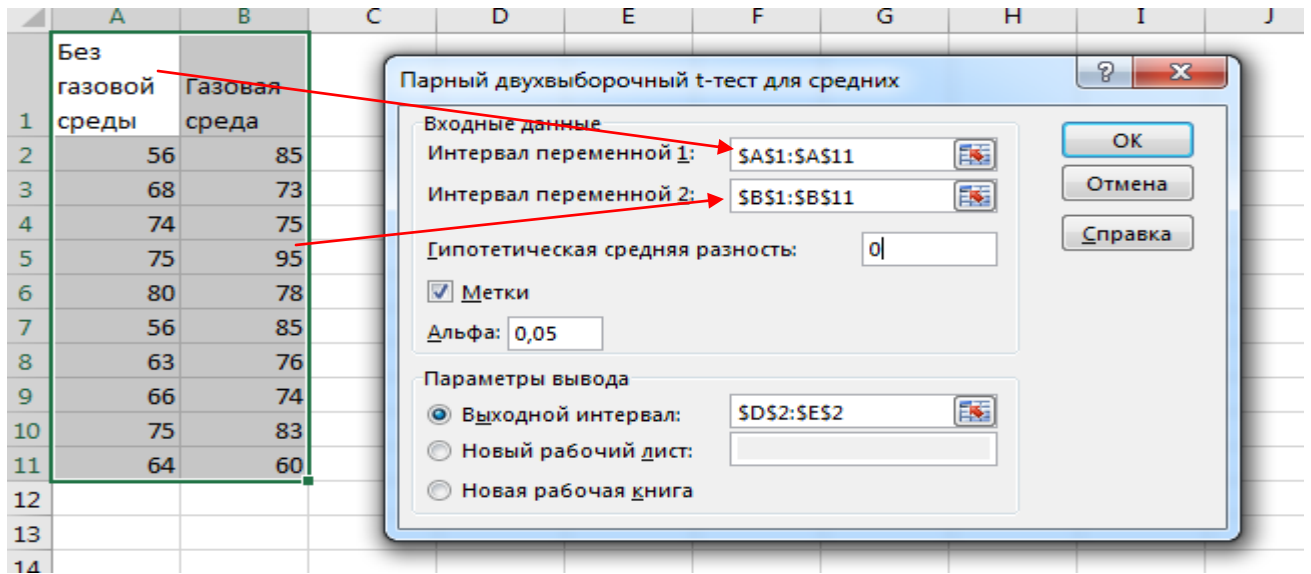


Рис. 3.16. Диалоговое окно **Парный двухвыборочный t-тест для средних**
После нажатия на клавишу **ОК** получаем итоговую таблицу (рис. 3.17).

	А	В	С	Д	Е	Ф	Г
1	Без газовой среды	Газовая среда					
2	56	85					
3	68	73					
4	74	75					
5	75	95					
6	80	78					
7	56	85					
8	63	76					
9	66	74					
10	75	83					
11	64	60					
12							
13							
14							
15							
16							

Рис. 3.17. Итоговая таблица **t-теста результатов хранения яблок**

Так как фактическое значение критерия Стьюдента по модулю (t -статистика = 2,84) в нашем примере больше табличного (t -критическое двухстороннее = 2,26 при числе степеней свободы $df=10-1=9$), нулевая гипотеза отвергается. С 95% ве-

роятностью можно предполагать, что сохранность плодов в пакетах с газовой средой существенно выше.

Оценка результатов:

Компетенции	Владение теорией	Владение практикой	Балл
Умеет выбрать лучшие варианты на основе проверки гипотез	+	+	2
Умеет на удовлетворительно	+-	+-	1
Не владеет	-	-	0

Дата _____ Подпись преподавателя _____

Работа 4. Оценка двух вариантов при качественной изменчивости признаков

Теоретическая часть:

В исследованиях по биотехнологии часто приходится иметь дело с *качественной изменчивостью* признаков. К качественным признакам относят такие признаки, которые не поддаются количественному измерению: разные культуры, разные виды болезней, наличие или отсутствие признаков, разная форма и окраска колонии грибов, расщепление гибридов и т.д.

При этом вместо измерения, какого либо показателя, как это имеет место при количественной изменчивости, при качественной изменчивости подсчитывают число объектов с тем или иным признаком: число эксплантов, образовавших каллус, число поврежденных и здоровых растений, число товарных продуктов и т.п..

Основными статистическими показателями качественной изменчивости являются: доля признака, показатель изменчивости, коэффициент вариации и ошибка выборочной доли.

Контрольные вопросы к защите:

<p>1. Перечислите статистические характеристики качественной изменчивости.</p>	<p>3. Расчет статистических показателей качественной изменчивости при числе групп более двух.</p>
--	---

2. Как вычисляется и что показывает доверительный интервал для доли признака в совокупности?	4. Как оценить значимость между выборочными долями?
--	---

Общая постановка задачи:

1. В соответствии с номером задания рассчитать статистические показатели качественной изменчивости.
2. Подготовить данные к статистическому тесту и провести расчеты необходимых критериев для оценки существенности различий.
3. Сделать правильное статистическое и биологическое заключение по эффективности изученных приемов.

Список индивидуальных данных:

Пример 1. В опыте проводили оценку двух сортов картофеля (сорт А – обычный сорт (вариант1) и сорт Б – трансгенный (вариант 2) на пораженность фитофторой. Для учета пораженности отобрали по 100 клубней картофеля каждого сорта ($N_1=100, N_2=100$), при этом численность пораженных клубней фитофторой для обычного сорта (вариант 1) составила 40, а для трансгенного (вариант 2) – 30 (задание 1):

Количество пораженных клубней картофеля
фитофторой у обычного и трансгенного сортов, шт

Таблица 11

№ задания	1	2	3	4	5	6	7	8
Обычный	40	50	90	80	70	80	78	70
Трансгенный	30	15	55	57	40	60	44	45
№ задания	9	10	11	12	13	14	15	16
Обычный	56	68	77	56	85	45	56	64
Трансгенный	13	56	23	10	45	10	20	48

Пример 2. В опыте изучали режимы индукции каллусогенеза у растений томата *in vitro*. Сравнивали эффективность каллусогенеза на питательной среде Мурасиге-Скуга (MS) (вариант 1) и питательной среде Гамборга (B5) (вариант 2) с добавлением 2 мг/л 2,4-дихлорфеноксиуксусной кислоты (в обоих вариантах). Общее количество эксплантов на каждый вариант – 50 шт. ($N_1=50, N_2=50$).

Количество образовавшихся каллусов, шт

Таблица 12

№ задания	17	18	19	20	21	22	23	24
MS	40	45	36	30	28	30	37	35
B5	16	16	26	60	20	40	99	46
№ задания	25	26	27	28	29	30	31	32
MS	37	38	48	42	37	36	30	25
B5	30	19	35	78	8	20	16	34

Статистическая обработка

Вариант 1			Вариант 2		
$n_1 =$	$n_2 =$	$N_1 =$	$n_1 =$	$n_2 =$	$N_2 =$
$p_1 = \frac{n_1}{N_1} =$			$p_2 = \frac{n_1}{N_2} =$		
$q_1 = 1 - p_1 =$			$q_2 = 1 - p_2 =$		
$S_1 = \sqrt{p_1 \cdot q_1} =$			$S_2 = \sqrt{p_2 \cdot q_2} =$		
$V_1 = \frac{S_1}{S_{\max}} \cdot 100 =$			$V_2 = \frac{S_2}{S_{\max}} \cdot 100 =$		
$S_{p_1} = \sqrt{\frac{p_1 \cdot q_1}{N_1}} =$			$S_{p_2} = \sqrt{\frac{p_2 \cdot q_2}{N_2}} =$		
$p_1 \pm t_{05} \cdot S_{p_1} =$			$p_2 \pm t_{05} \cdot S_{p_2} =$		

Выбор лучшего варианта на основе:

а) критерия существенности (t):

$$t_{\phi} = \frac{|p_2 - p_1|}{\sqrt{S_{p_1}^2 + S_{p_2}^2}} = \frac{d}{S_d} \quad t_{05} =$$

б) доверительного интервала для генеральной доли:

$$p_1 \pm t_{05} \cdot S_{p_1} =$$

$$p_2 \pm t_{05} \cdot S_{p_2} =$$

в) доверительного интервала для генеральной разности долей:

$$d \pm t_{05} \cdot S_d =$$

Вывод:

Оценка результатов:

Компетенции	Владение теорией	Владение практикой	Балл
Освоил оценку вариантов по качественным признакам на хорошо	+	+	2
Освоил на удовлетворительно	+-	+-	1
Не владеет	-	-	0

Дата _____ Подпись преподавателя _____

Работа 5. Оценка соответствия между фактическими и теоретическими наблюдениями

Теоретическая часть:

В исследовательской работе порой возникает задача сопоставления фактических (эмпирических) частот с теоретическими (ожидаемыми) частотами с тем, чтобы установить, значимы или случайны различия между ними. Нулевая гипотеза сводится к предположению, что несоответствие эмпирических частот, вычисленных по тому или иному закону распределения – совершенно случайное, т.е. между вычисленными и эмпирическими частотами никакой разницы нет. Для проверки нулевой гипотезы, когда необходимо определить соответствие двух сравниваемых рядов распределения – эмпирического и теоретического или двух эмпирических или гипотезы о независимости 2 или нескольких признаков, наиболее часто применяется критерий χ^2 – Хи-квадрат (критерий Пирсона).

Применение критерия χ^2 требует определенной осторожности: критерий χ^2 используется при изучении качественных признаков, поэтому в формуле должны подставляться только частоты, а не величины, полученные измерением, взвешиванием и т.д. При проверке гипотезы о соответствии эмпирических распределений нормальному распределению, желательно иметь не менее 50 наблюдений, а в каждой теоретически рассчитанной группе – не менее пяти наблюдений. Если крайние группы в ряду распределения малочисленны, их необходимо объединить.

Контрольные вопросы к защите:

1. Условия применения критерия χ^2 – Хи-квадрат	2. Примеры использования критерия χ^2 – Хи-квадрат
--	---

Общая постановка задачи:

1. Уясните сущность критерия χ^2 – Хи-квадрат.
2. Рассчитайте теоретические частоты и проверьте нулевую гипотезу на соответствие фактического распределения частот (численностей) теоретическому или независимости признаков и сделайте вывод.

1. Проверка гипотезы на независимость:

Пример 1. Изучали действие жидкого биопрепарата *Псевдобактерин* – 2, полученного на основе штамма *Pseudomonas sp. BS 1393* на заболевание картофеля паршой. Результаты представлены в таблице 13.

Таблица 13

Варианты	Число клубней, пораженных паршой								Сумма
	не пораж.		слабо		средне		сильно		
	<i>f</i>	<i>F</i>	<i>f</i>	<i>F</i>	<i>f</i>	<i>F</i>	<i>f</i>	<i>F</i>	
Без обработки	5		20		10		15		50
Обработка биопрепаратом	9		23		14		4		50
Сумма	14	14	43	43	24	24	19	19	100

Пример 2. В хранилище изучали влияние газовой среды №1 (12% O₂ и 9% CO₂) и газовой среды №2 (3% O₂ и 5% CO₂) на сохранность плодов груши. Результаты эксперимента представлены в таблице 13:

Таблица 12

Тип газовой среды	Количество плодов				Сумма
	товарные		нетоварные		
	<i>f</i>	<i>F</i>	<i>f</i>	<i>F</i>	
Среда №1	135		65		200
Среда №2	154		26		180
Сумма	289	289	91	91	380

Алгоритм теста:

1. Исходя, из нулевой гипотезы о независимости действия изучаемых признаков рассчитываем для всех примеров теоретические (ожидаемые) частоты и записываем их в таблицу.

$$F_{ij} = \frac{\sum f_i \cdot \sum f_j}{\sum f_{ij}} =$$

где: F_{ij} – ожидаемая частота для i -й строки и для j -го столбца

$\sum f_i$ – сумма фактических частот для i -й строки

$\sum f_j$ – сумма фактических частот для j -го столбца

$\sum f_{ij}$ – общая сумма всех наблюдений

2. Рассчитываем отклонения фактических частот от ожидаемых и возводим их в квадрат по формуле: $d^2 = (f_{ij} - F_{ij})^2 =$

3. Находим сумму отношений квадратов отклонений частот к теоретически ожидаемой частоте и эту сумму отношений принимаем за критерий Хи-квадрат фактический – $\chi^2_{ф}$

$$\chi^2_{ф} = \sum \frac{(f - F)^2}{F} = \sum \frac{d^2}{F} =$$

4. Находим табличное значение χ^2 и делаем вывод.

$$\chi^2_{05} = \quad \text{при } df = (c - 1)(k - 1) =$$

$$\chi^2_{01} = \quad c - \text{число строк, } k - \text{число колонок}$$

Вывод:

Оценка соответствия между фактическими и теоретическими наблюдениями в программе Excel

Изучали действие жидкого биопрепарата *Псевдобактерин – 2*, полученного на основе штамма *Pseudomonas sp. BS 1393* на заболевание картофеля паршой.

Варианты	Число клубней, пораженных паршой						Сумма
	слабо		средне		сильно		
	f	F	f	F	f	F	
Без обработки	30		15		5		50
Обработка биопрепаратом	23		12		15		50
Сумма	53	53	27	27	20	20	100

Выдвигаем нулевую гипотезу H_0 : пораженность клубней картофеля паршой не зависит от обработки биопрепаратом *Псевдобактерин – 2*.

Альтернативная гипотеза H_a : пораженность клубней картофеля паршой зависит от обработки биопрепаратом *Псевдобактерин – 2*.

Проверку гипотезы на независимость пораженности клубней картофеля паршой от обработки биопрепаратом *Псевдобактерин – 2* в программе Excel можно провести следующим образом:

1. В активном листе программы Excel построим таблицу с фактическим количеством пораженных клубней (рис.5.1)

2. В этом же листе построим таблицу для ожидаемых (теоретических) частот с пустыми ячейками (рис. 5.1).

Таблица фактических частот					Таблица ожидаемых (теоретических) частот				
Варианты	Число клубней, пораженных паршой			Сумма	Варианты	Число клубней, пораженных паршой			Сумма
	слабо	средне	сильно			слабо	средне	сильно	
Без обработки	30	15	5	50	Без обработки				
Обработка биопрепаратом	23	12	15	50	Обработка биопрепаратом				
Сумма	53	27	20	100	Сумма				

Рис. 5.1. Таблицы фактических и ожидаемых частот

3. Для вычисления ожидаемых частот и их автоматического заполнения воспользуемся функцией массива.

4. Правой клавишей мышки выделим пустые ячейки в таблице ожидаемых (теоретических) частот размером 2x3, далее поставим знак =, выделим строку с суммами по столбцам **B7:D7**, умножим на столбец с суммами по строкам **E5:E6**, разделим на общую сумму **E7** и далее нажмем одновременно на сочетание клавиш **Ctrl + Shift + Enter**. В строке формул появится функция массива – в фигурных скобках $\{=B7:D7*E5:E6/E7\}$

5. После нажатия на указанные клавиши в таблице автоматически появляются значения ожидаемых или теоретических частот (рис. 5.2).

Таблица фактических частот					Таблица ожидаемых (теоретических) частот				
Варианты	Число клубней, пораженных паршой			Сумма	Варианты	Число клубней, пораженных паршой			Сумма
	слабо	средне	сильно			слабо	средне	сильно	
Без обработки	30	15	5	50	Без обработки	26,5	13,5	10	
Обработка биопрепаратом	23	12	15	50	Обработка биопрепаратом	26,5	13,5	10	
Сумма	53	27	20	100	Сумма				

Рис. 5.2. Расчет ожидаемых частот функцией массива

6. Для проверки нулевой гипотезы на независимость в контекстном меню функций программы Excel выберем категорию **Статистические**. В появившемся перечне конкретных статистических функций выберем **ХИ2-ТЕСТ** (рис. 5.3).

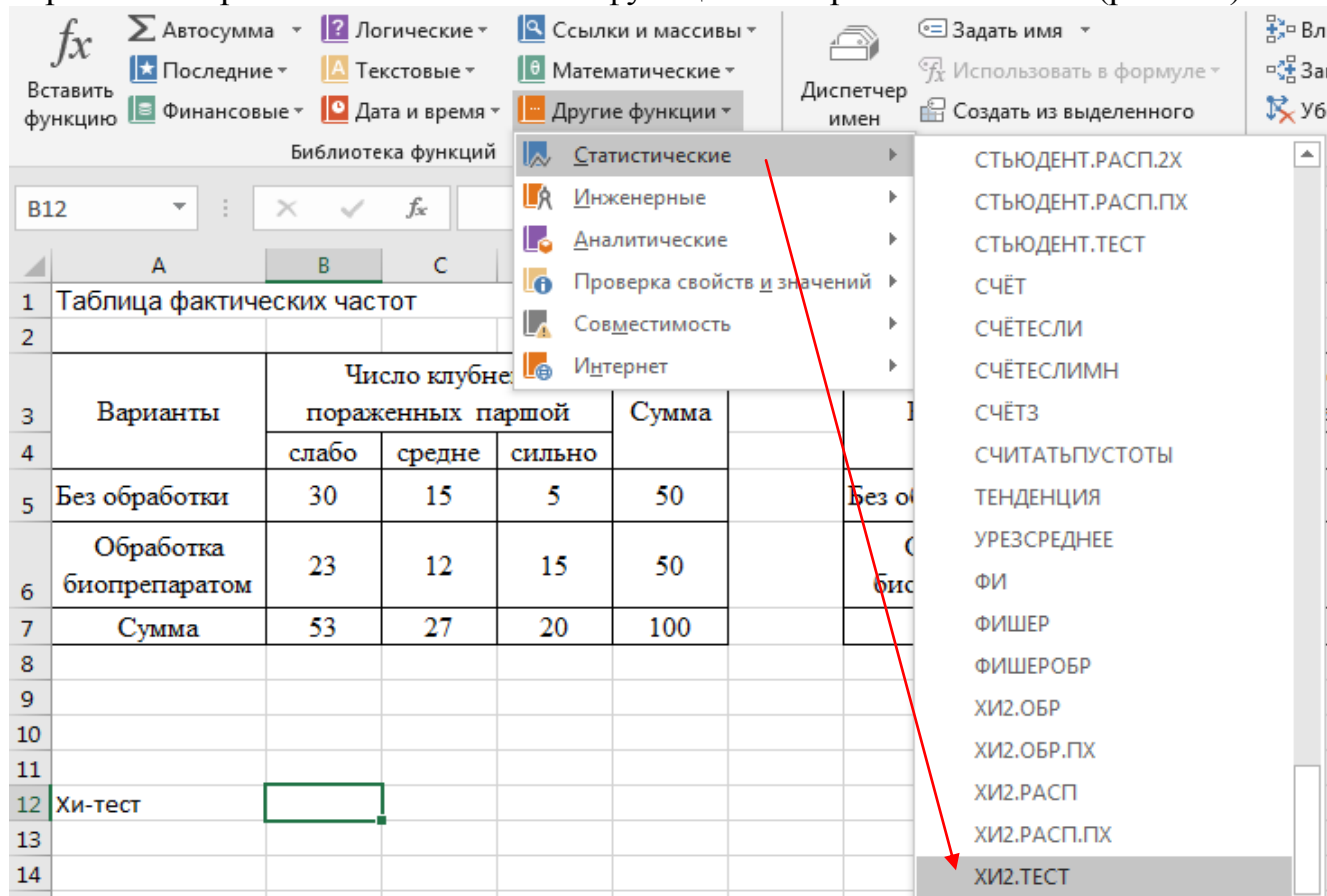


Рис. 5.3. Контекстное меню выбора статистических функций

7. Далее открывается окно для выбора аргументов функции (рис. 5.4). В поле **Фактический интервал** ставим курсор и мышкой выбираем диапазон ячеек из таблицы значений фактических частот **B5:D6**, в поле **Ожидаемый интервал** ставим курсор и мышкой выбираем диапазон ячеек из таблицы значений ожидаемых

частот **H5:J6**. После нажатия на клавишу **OK** получаем значение **ХИ2 ТЕСТ = 0,04376565**. (рис. 5.4.)

ХИ2ТЕСТ =ХИ2ТЕСТ(В5:D6;H5:J6)

Расчет Хиквадрат

Варианты	Число клубней, пораженных паршой			Сумма	Варианты	Число клубней, пораженных паршой			Сумма
	слабо	средне	сильно			слабо	средне	сильно	
Без обработки	30	15	5	50	Без обработки	26,5	13,5	10	
Обработка биопрепаратом	23	12	15	50	Обработка биопрепаратом	26,5	13,5	10	
Сумма	53	27	20	100	Сумма				

9 Хи-тест =ХИ2ТЕСТ(В5:J6)

Аргументы функции

ХИ2ТЕСТ

Фактический_интервал В5:D6 = {30; 15; 5; 23; 12; 15}

Ожидаемый_интервал H5:J6 = {26,5; 13,5; 10; 26,5; 13,5; 10}

= 0,043764565

Возвращает тест на независимость: значение распределения Хи-квадрат для статистического распределения и соответствующего числа степеней свободы.

Ожидаемый_интервал диапазон, содержащий отношение произведений итогов по строкам и столбцам к общему итогу.

Значение: 0,043764565

[Справка по этой функции](#)

Рис. 5.4. Диалоговое окно для выбора аргументов функции ХИ2-ТЕСТ

8. Полученное значение **ХИ2-ТЕСТ**- 0,04376565 является вероятностью (p) для отклонения нулевой гипотезы. Так как $p = 0,04376565$ меньше вероятности принятого уровня значимости $p = 0,05$, с вероятностью 95% отвергаем гипотезу о независимости пораженности клубней картофеля паршой от обработки биопрепаратом *Псевдобактерин*.

9. Фактическое значение критерия χ^2 – Хи-квадрат ($\chi^2_{\text{фак}}$) можно рассчитать, если из контекстного меню статистических функций выбрать функцию **ХИ2ОБР** (рис. 5.3). В появившемся диалоговом окне (рис. 5.5.) в поле **Вероятность** введем значение 0,04376565, в поле **Степени свободы** введем цифру 2. Для нашего примера таблицы 2×3 число строк (число вариантов) равно 2, число колонок (число групп) равно 3. Число степеней свободы $df = (c - 1)(k - 1) = (2 - 1)(3 - 1) = 2$.

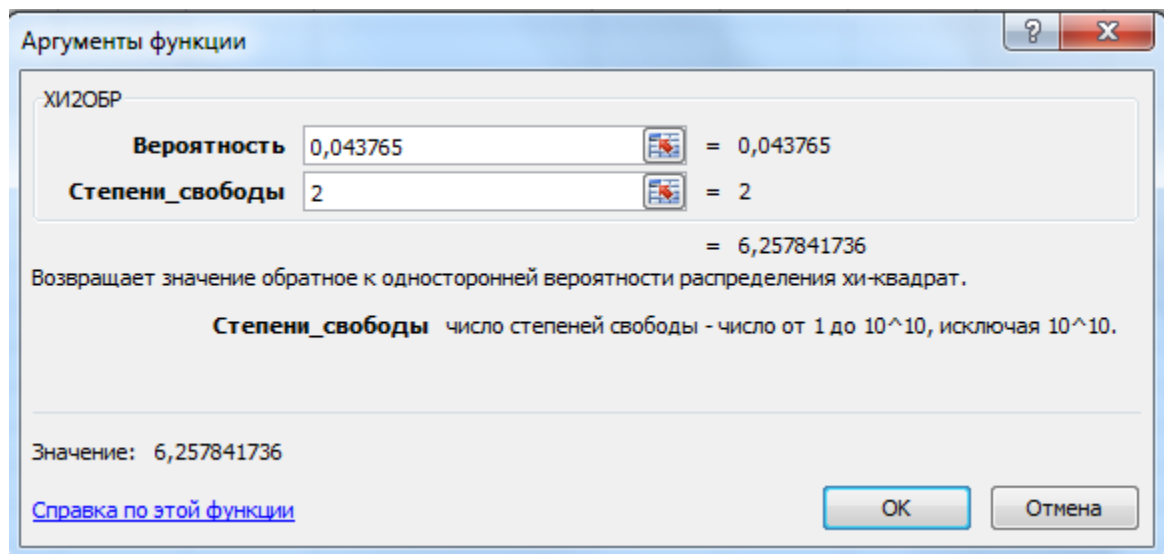


Рис. 5.5. Диалоговое окно для расчета фактического значения Хи-квадрат

10. После нажатия на клавишу **ОК** получаем фактическое значение критерия $\chi_{\text{фак}}^2 = 6,2578$ (рис.5.5). Табличное значение критерия $\chi_{05}^2 = 5,99$ при числе степеней свободы $df = 2$. Так как $\chi_{\text{фак}}^2 > \chi_{05}^2$ нулевая гипотеза о независимости пораженности клубней картофеля паршой от применения от обработки биопрепаратом *Псевдобактерин* с вероятностью ошибки 5% отвергается. Таким образом, мы приходим к аналогичному выводу, как и при проверке нулевой гипотезы по значению вероятности (p) ХИ2ТЕСТ.

2. Проверка гипотезы соответствия (согласия).

а) Проверка соответствия наблюдаемых распределений случайной переменной какому-либо теоретическому распределению или закону. H_0 и H_A записываются следующим образом:

$$H_0: F(x) = F_0(x) \text{ и } H_A: F(x) \neq F_0(x),$$

где $F(x)$ — функция (эмпирическая кривая) распределения значений, полученная на основе выборки, а $F_0(x)$ — теоретическая кривая.

б) Оценка соответствия вероятности события (исхода) P ожидаемой (теоретической) величине (P_0), т.е. $H_0: P = P_0$ $H_A: P \neq P_0$.

Пример 1. В опытах по скрещиванию сортов гороха изучали распределение формы (гладкая, морщинистая) и окраски (жёлтая, зелёная) плодов у гибридов. Согласно установленным закономерностям (закон Менделя), распределение фенотипов: гладкие + жёлтые, гладкие + зелёные, морщинистые + жёлтые и морщинистые + зелёные — должно отвечать соотношению 9:3:3:1. Любая форма и окраска плодов является случайным исходом.

При дигибридном скрещивании во втором поколении произошло расщепление $n=381$ горошин на: гладкие и жёлтые ($f_1 = 221$), гладкие и зелёные ($f_2 = 67$), морщинистые и жёлтые ($f_3 = 74$), морщинистые и зелёные ($f_4 = 19$). Соответствует ли наблюдаемое расщепление закону Менделя?

1. Исходя из закона Менделя рассчитаем вероятности (P по H_0) отдельных фенотипов составят: гладко-жёлтые $P_1 = 9/16$, гладко-зелёные $P_2 = 3/16$, морщинисто-жёлтые $P_3 = 3/16$ и морщинисто-зелёные $P_4 = 1/16$.

2. Расчет ожидаемых (теоретических) численностей в соответствии с H_0 :

$$F_1 = P_1 \cdot n =$$

$$F_2 = P_2 \cdot n =$$

$$F_3 = P_3 \cdot n =$$

$$F_4 = P_4 \cdot n =$$

$$\chi^2_{\phi} = \sum \frac{(f - F)^2}{F} = \sum \frac{d^2}{F} =$$

3. Находим табличное значение χ^2 и делаем вывод.

$$\chi^2_{05} =$$

$$\text{при } df = (k - 1) =$$

$$\chi^2_{01} =$$

k – число групп

Вывод:

Оценка результатов:

Компетенции	Владение теорией	Владение практикой	Балл
Владеет методами оценки соответствия между эмпирическими и ожидаемыми наблюдениями	+	+	2
Владеет на удовлетворительно	+-	+-	1
Не владеет	-	-	0

Дата _____ Подпись преподавателя _____

Работа 6. Обработка данных с применением непараметрических критериев

1. Критерий Ван-дер-Вандена для сравнения независимых выборок.

Пример 1. В лабораторных условиях проводилась оценка устойчивости 2-х сортов овса к стеблевой ржавчине. Растения опрыскивали из пульверизатора суспензией уредоспор. На 14-й день после инокуляции учитывали тип реакции в баллах по Стэкмену и Левину.

Таблица 14

Сорта	Номера проб (растений)									
	1	2	3	4	5	6	7	8	9	10
А	2	2	1	4	3	3	4	3	2	3
Б	1	0	2	1	0	0	3	3	2	1

Решение:

1. Составляем вспомогательную таблицу, в которой ранжируем (располагаем в порядке возрастания) обе выборки в один вариационный ряд и каждому значению признака присваиваем порядковый номер (ранг) R .

2. Сумма всех наблюдений $n = n_1 + n_2 =$

Таблица 15

Пораженность в баллах		Ранг, R	$\frac{R}{n+1}$	$\psi = \left[\frac{R}{n+1} \right]$
Сорт А	Сорт Б			
$n_1 =$	$n_2 =$	$n_1 + n_2 =$		$\sum \psi = \left[\frac{R}{n+1} \right] =$

3. Находим для всех значений признака одной из выборок отношение $\frac{R}{n+1}$

4. По таблице приложений 4 определяем значение функции (пси) $\psi = \left[\frac{R}{n+1} \right]$

5. Суммируем с учетом знаков значения функции (пси) ψ и эту сумму принимаем за фактический критерий Ван-дер-Вандена X .

6. По таблице приложений 5 находим табличное значение X при $n = n_1 + n_2$, с учетом разности $n_1 - n_2$ и сравниваем фактическое значение X_ϕ с табличным X_{05}

Вывод:

2. Критерий Уайта для сравнения независимых выборок.

Пример 2. Изучено влияние обработки семян биопрепаратом Бионаноплант на всхожесть семян озимой пшеницы, %

Таблица 16

Варианты	Номера делянок									
	1	2	3	4	5	6	7	8	9	10
Контроль	82	86	81	76	85	89	91	90	75	75
Бионаноплант	94	92	88	92	88	95	90	90	97	–

Решение:

1. Располагаем все значения сравниваемых выборок в один ранжированный ряд, и каждому значению признака присваиваем ранг, причем одинаковые значения признаков имеют один и тот же средний ранг.

%	
Ранг	

2. Определяем сумму рангов по каждому сорту: $\sum R_1 = \quad \sum R_2 =$

3. Меньшая сумма рангов принимается за T -критерий Уайта $\sum R_{\min} = T_\phi =$

4. По таблице приложений 6 находим табличное значение T -критерия при $n_1 = \quad ; n_2 = \quad ;$ и сравниваем T_ϕ с T_{05}

Вывод:

3. Критерий Вилкоксона для сравнения зависимых выборок.

Пример 3. 10 корзинок подсолнечника перед цветением делили условно пополам. Одну половину корзинок опыляли пылью своих же цветков, а на вторую половину перед самоопылением наносили в избытке пыльцу кукурузы. Эффект дополнительного опыления определяли по числу завязавшихся семян.

Таблица 17

Число семян подсолнечника при 2-х способах опыления

Способ опыления	№№ растений									
	1	2	3	4	5	6	7	8	9	10
самоопыл	11	30	18	20	39	0	94	34	16	25
то же+ кукуруза	26	37	14	22	64	15	100	34	15	28
d										
R										

Решение:

1. Определяем отклонения по всем сопряженным парам $d_i = X_i^A - X_i^B$

2. Присваиваем ранг отклонениям (d) без учета знаков, при этом нулевые отклонения рангов не имеют.

3. Определяем отдельно сумму рангов для отрицательных и положительных отклонений $\sum R_1 =$; $\sum R_2 =$

4. Меньшую сумму рангов принимаем за критерий Вилкоксона $\sum R_{\min} = W_\phi =$

5. По таблице приложений 7 определяем теоретическое значение $W_{05} =$
 при: $\gamma = n - k =$, где k – число нулевых эффектов. При $W_\phi \geq W_{05}$ $H_0: d=0$

Вывод:

4. Критерий (Z) для сравнения зависимых выборок.

Пример 4. Поражение листьев яблони при инокуляции штаммами *Venturia inaequalis* (парша) определялось по диаметру пятна, мм:

Таблица 18

Штамм	№ листа											
	1	2	3	4	5	6	7	8	9	10	11	12
№1	0	0	5	4	3	4	4	5	6	3	3	4
№2	0	0	7	6	5	3	7	8	8	5	3	6
±												

1. Отклонения (эффекты) обозначаем знаками «-» и «+», затем суммируем число этих эффектов, причем нулевые разности в расчет не принимаются.

2. Сумму положительных эффектов принимаем за критерий Z_{ϕ}

3. По таблице приложений 8 определяем Z_{05} при $n_1 = n - k$

Вывод:

Оценка результатов:

Компетенции	Владение теорией	Владение практикой	Балл
Владеет методами оценки соответствия между эмпирическими и ожидаемыми наблюдениями	+	+	2
Владеет на удовлетворительно	+-	+-	1
Не владеет	-	-	0

Дата _____ Подпись преподавателя _____

Работа 7. Корреляционно-регрессионный анализ в биотехнологических исследованиях

Теоретическая часть:

В биологических исследованиях редко приходится иметь дело с точными и определенными функциональными связями, когда каждому значению независимого признака (X) соответствует строго определенное значение результативного (зависимого) признака (Y). Чаще между изучаемыми явлениями, объектами, условиями среды, ростом, продуктивностью растений и другими показателями существуют корреляционные или вероятностные взаимосвязи, когда определенному значению независимой переменной X соответствует не одно, а множество возможных значений признака Y .

Корреляции подразделяют по направлению, форме и числу связей. По направлению корреляция может быть прямой или обратной, по форме прямолинейной и криволинейной, а по числу связей – простой и множественной.

На основании корреляционного анализа устанавливают форму, направление и тесноту связи, то есть дается качественная оценка зависимости.

Для количественной оценки связи между изучаемыми признаками проводится регрессионный анализ, на основании которого определяют уравнение регрессии и строят теоретическую линию регрессии.

Контрольные вопросы к защите:

1. Виды корреляции.	3. Сущность регрессии и способы ее определения.
2. Чем и как измеряется сила и направление связи?	4. Как использовать результаты регрессионного анализа для прогнозирования.

Общая постановка задачи:

1. Усвоить сущность и значение корреляционного и регрессионного анализов, условия и место их применения в опытном деле.
2. Уяснить принципиальную разность между корреляционной и функциональной, между прямолинейной и криволинейной связями.
3. Освоить технику корреляционного и регрессионного анализов.
4. Сделать статистический и биологический вывод по характеру связей изучаемых признаков.

А. Линейная корреляция и регрессия

Список индивидуальных данных:

Таблица 19

Пример 1. Масса зерна ячменя (X , мг) и содержание жира в зерне (Y ,%)

1 задание		2 задание		3 задание		4 задание		5 задание	
Y	X	Y	X	Y	X	Y	X	Y	X
1,2	11,0	2,1	13,5	1,1	10,5	1,1	12,0	1,0	10,7
5,1	19,9	1,0	10,8	5,0	19,8	5,1	21,9	3,1	16,8
2,3	15,9	2,3	15,6	2,4	15,0	1,2	11,0	1,5	14,0
3,1	16,3	3,6	17,5	3,3	19,6	2,2	16,9	2,0	17,1
0,9	10,2	4,1	18,4	2,8	16,7	4,3	19,4	5,2	25,3
4,1	21,4	2,7	16,0	3,0	18,3	2,2	17,3	4,1	20,4
2,1	15,8	4,7	21,0	2,2	16,8	3,1	18,8	2,1	15,4
4,2	21,6	5,2	24,9	1,7	13,5	2,2	15,9	4,1	23,1
1,1	12,3	1,0	10,5	2,5	15,7	2,1	16,0	3,1	18,9
3,4	17,3	2,8	17,2	3,3	19	2,9	18,9	2,7	17,6

Таблица 20

Пример 2. Зависимость антиоксидантной активности биомассы зеленых культур (Y , $1/EC_{50}$) от концентрации фенолов (X , мг/г сухой массы)

6 задание		7 задание		8 задание		9 задание		10 задание	
Y	X	Y	X	Y	X	Y	X	Y	X
0,15	45	0,03	12	0,03	10	0,01	5	0,05	15
0,13	42	0,06	25	0,04	15	0,13	32	0,03	10
0,08	23	0,02	10	0,07	20	0,16	45	0,07	20
0,09	15	0,16	50	0,06	21	0,03	10	0,01	5
0,08	10	0,06	20	0,13	45	0,05	15	0,06	20
0,10	23	0,05	15	0,14	48	0,05	14	0,07	30
0,06	14	0,03	10	0,10	30	0,04	8	0,08	35
0,05	10	0,01	5	0,02	5	0,10	20	0,16	50
0,05	8	0,12	35	0,10	25	0,12	25	0,11	45
0,06	15	0,03	11	0,12	35	0,13	30	0,09	25

Пример 3. На агаризированных питательных средах в чашках Петри изучали зависимость роста колоний гриба *Fusarium oxysporum var.lini* (Y , диаметр грибных колоний, мм) от времени (X , час). Измерение диаметра грибных колоний проводили каждые 12 часов.

Таблица 21

11 задание		12 задание		13 задание		14 задание		15 задание	
Y	X	Y	X	Y	X	Y	X	Y	X
0	12	2	12	1	12	2	12	1	12
5	24	6	24	5	24	6	24	5	24
8	36	9	36	10	36	11	36	10	36
12	48	11	48	13	48	13	48	12	48
20	60	15	60	17	60	22	60	20	60
24	72	23	72	25	72	25	72	23	72
28	84	31	84	26	84	30	84	33	84
30	96	33	96	30	96	33	96	39	96
32	108	40	108	39	108	40	108	41	108
45	120	48	120	41	120	46	120	50	120

Таблица 22

Пример 4. Зависимость урожайности озимой пшеницы (Y , ц/га) от пораженности бурой ржавчины (X , %)

16 задание		17 задание		18 задание		19 задание		20 задание	
Y	X	Y	X	Y	X	Y	X	Y	X
51,1	20,2	49,3	36,1	50,3	22,3	49,7	21,7	55,1	9,9
50,3	27,4	52,5	22,4	50,4	28,0	50,2	24,4	53,0	25,4
51,1	21,6	55,3	8,5	43,4	54,1	52,5	15,3	50,2	21,1
48,4	46,0	43,8	54,1	46,5	45,3	48,1	36,2	51,1	35,6
49,0	41,4	48,1	43,2	55,3	10,2	41,0	59,7	50,7	42,7
49,1	43,3	39,6	64,1	50,3	20,0	53,0	25,8	51,5	34,0
48,5	37,8	41,3	57,0	40,2	60,4	51,5	34,2	45,8	54,2
49,0	42,5	55,3	12,4	55,3	10,1	45,2	54,6	50,5	43,6
41,8	59,1	48,0	43,1	43,8	54,8	42,0	54,0	58,8	12,4
45,0	54,7	40,6	56,3	48,5	36,2	41,6	59,6	40,2	60,1

Вычислите коэффициенты линейной корреляции и регрессии, рассчитайте уравнение регрессии и представьте корреляционную зависимость на графике.

Решение:

Признаки		Квадраты		Произведения
Y	X	Y ²	X ²	Y·X
Суммы				

Число пар сравнений n =
 Средние по ряду Y и по ряду X

$$\bar{y} = \sum Y : n =$$

$$\bar{x} = \sum X : n =$$

$$\sum (Y - \bar{y})^2 = \sum Y^2 - \frac{(\sum Y)^2}{n} =$$

$$\sum (X - \bar{x})^2 = \sum X^2 - \frac{(\sum X)^2}{n} =$$

$$\sum (X - \bar{x}) \cdot (Y - \bar{y}) = \sum XY - \frac{\sum X \cdot \sum Y}{n} =$$

$$r = \frac{\sum (X - \bar{x}) \cdot (Y - \bar{y})}{\sqrt{\sum (X - \bar{x})^2 \cdot \sum (Y - \bar{y})^2}} =$$

$$S_r = \sqrt{\frac{1 - r^2}{n - 2}} =$$

$$d_{yx} = r^2 =$$

$$b_{yx} = \frac{\sum (X - \bar{x}) \cdot (Y - \bar{y})}{\sum (X - \bar{x})^2} =$$

$$S_{b_{yx}} = S_r \cdot \sqrt{\frac{\sum (Y - \bar{y})^2}{\sum (X - \bar{x})^2}} =$$

$$S_{yx} = S_r \cdot \sqrt{\sum (Y - \bar{y})^2} =$$

$$t_r = t_b = \frac{r}{s_r} =$$

$$t_{0.5} = \quad \text{при } df = n - 2 =$$

Определяем значения Y для экстремальных значений X (X_{min} и X_{max}) и строим теоретическую линию регрессии Y по X по уравнению $Y = \bar{y} + b_{yx} \cdot (X - \bar{x}) =$

Определяем по уравнению значения Y для экстремальных значений X (X_{min} и X_{max}) и строим теоретическую линию регрессии Y по X .

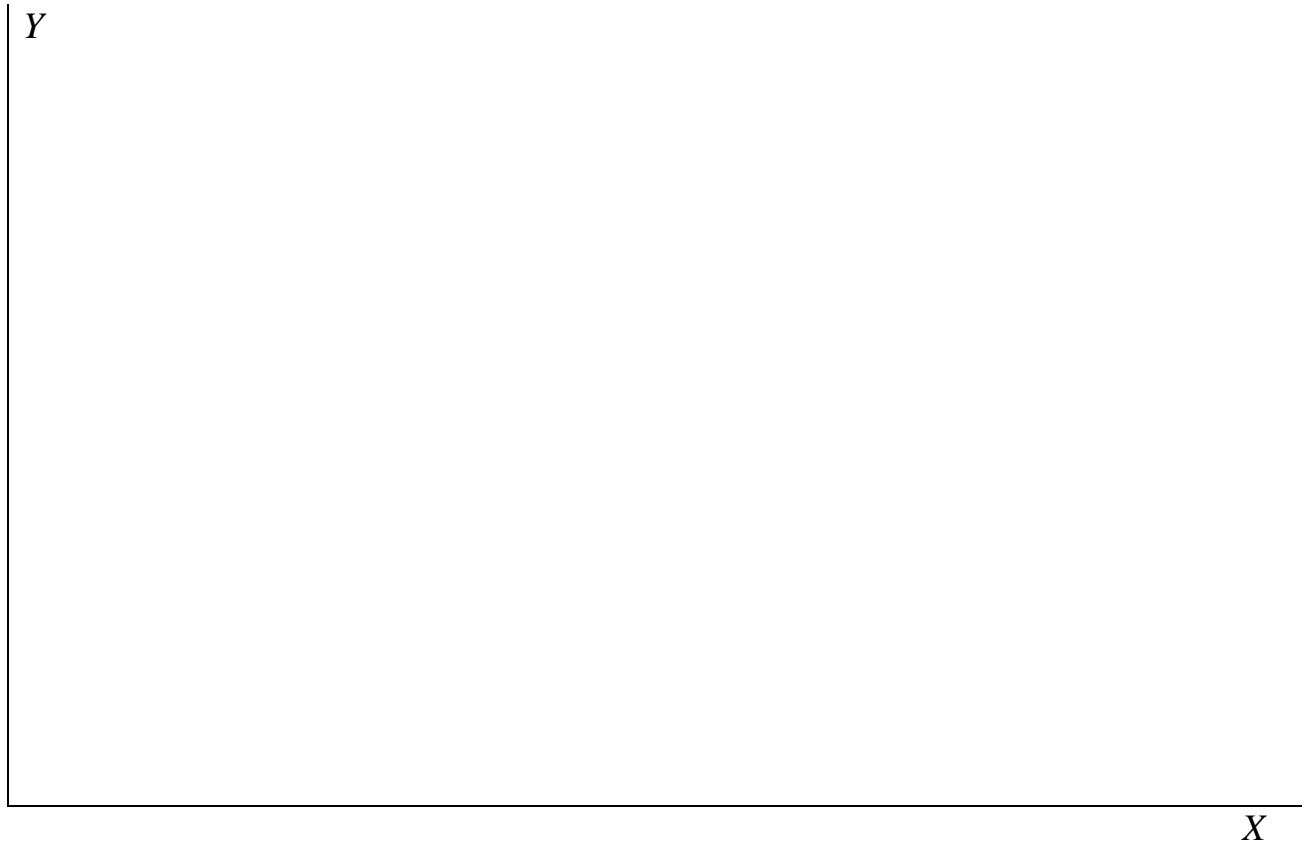


Рис. График прямолинейной корреляционной зависимости

Вывод:

В. Нелинейная корреляция и регрессия

Если связь между изучаемыми признаками носит нелинейный характер, что можно легко установить на графике по корреляционному полю, коэффициент прямолинейной корреляции непригоден в качестве меры связи. Он может указать на отсутствие сопряженности там, где налицо сильная криволинейная зависимость. Поэтому необходим новый показатель, который правильно измерял бы степень криволинейной зависимости. Таким показателем является корреляционное отношение, обозначаемое греческой буквой η (эта). Оно измеряет степень корреляции при любой ее форме. По корреляционному полю на графике можно определить характер связи (прямолинейная или криволинейная), а также тип аппроксимации. Графическое изображение изучаемой зависимости является очень удобным визуальным способом оценки адекватности регрессионной модели.

Пример 1. Степень повреждения побегов черной смородины второго поколения листовой галлицей $Y(\%)$ в зависимости от количества эфирных железок на листе X (шт).

X	48	50	53	54	57	61	61	61	62	56
Y	58	72	58	40	68	28	45	28	52	54
X	69	72	98	112	118	119	126	188	165	120
Y	58	25	22	5	15	2	5	5	3	2

В соответствии с номером задания вычислите коэффициент криволинейной корреляции (корреляционное отношение), определите его существенность и представьте графическую зависимость

Алгоритм расчетов:

1. Составим вспомогательную таблицу, в которой все значения зависимого признака X расположим по ранжиру в возрастающем порядке.
2. Разобьем весь ряд (колонку) признака X на 4 – 8 групп, так чтобы в каждой группе было не менее 2-х значений.
3. По каждой группе определим среднюю групповую \bar{x}_y .
4. Определим общую среднюю по ряду X (\bar{x}).
5. Для каждой группы ряда \bar{x}_y определим среднюю групповую по ряду Y (\bar{y}_x).
6. Определим общую среднюю по ряду Y (\bar{y}).
7. Определим суммы квадратов отклонений для группового варьирования признака $Y \sum (Y - \bar{y}_x)^2$ и общего $\sum (Y - \bar{y})^2$

Вспомогательная таблица

X	\bar{x}_y	Y	\bar{y}_x	$Y - \bar{y}_x$	$Y - \bar{y}$	$(Y - \bar{y}_x)^2$	$(Y - \bar{y})^2$
Суммы							

8. Рассчитаем:

корреляционное отношение

$$\eta_{yx} = \sqrt{\frac{\sum(Y - \bar{y})^2 - \sum(Y - \bar{y}_x)^2}{\sum(Y - \bar{y})^2}} =$$

ошибку корреляционного отношения

$$S_{\eta_{yx}} = \sqrt{\frac{1 - \eta_{yx}^2}{n - 1}} =$$

критерий существенности

$$t_{\eta_{yx}} = \frac{\eta_{yx}}{S_{\eta_{yx}}} =$$

$$t_{05} = \quad \text{при } df = n - 2$$

9. Строим график корреляционной зависимости 2-х признаков.

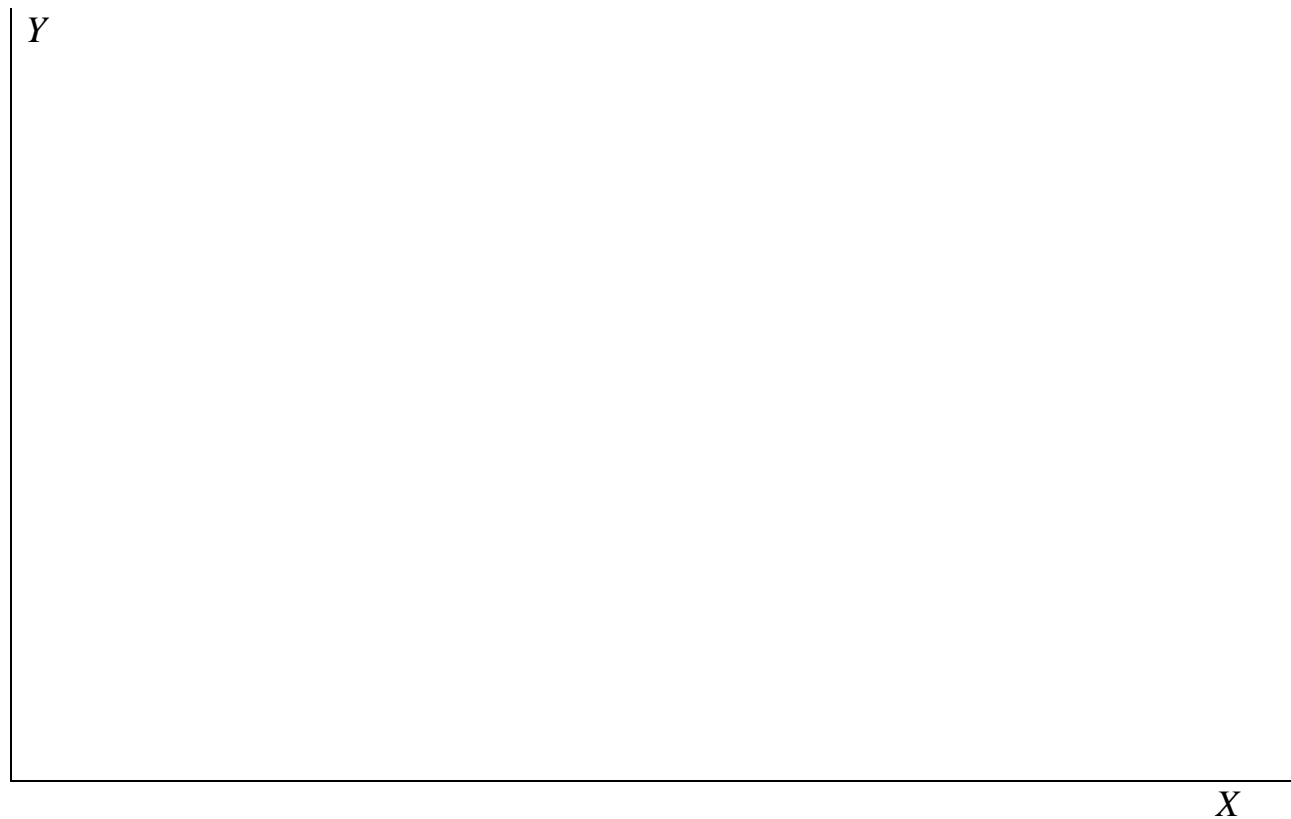


Рис. График нелинейной корреляционной зависимости
Вывод:

С. Ранговая корреляция

Применение коэффициента корреляции Пирсона r для оценки степени тесноты связи между признаками, особенно в той части, которая связана с оценкой его существенности, является обоснованным лишь в условиях нормального или близкого к нормальному распределению признаков в изучаемой совокупности. Если обе переменные или одна из переменных не подчиняются нормальному распределению, например, данные выражены в порядковых или номинальных шкалах (условные единицы, баллы), для установления сопряженности между признаками рекомендуется использовать коэффициент ранговой корреляции Спирмана – r_s

Коэффициенты корреляции рангов могут изменяться в пределах от -1 до $+1$, как и линейный коэффициент корреляции.

Пример. Определить корреляционную зависимость между степенью развития септориоза (X) и урожайностью яровой пшеницы (Y).

Таблица 24

X, балл	Y, ц/га	Ранги рядов		$d=R_x-R_y$	d^2
		R_x	R_y		
0	45				
1	42				
2	43				
3	39				
1	40				
4	38				
5	38				
2	44				
1	45				
5	36				
				$\Sigma d^2 =$	

Порядок выполнения работы:

1. Ранжируем отдельно значения признаков по ряду X и ряду Y. Одинаковым значениям признаков присваивают один и тот же средний ранг.

2. Определяем отклонения между рангами сопряженных значений признаков X и Y по формуле $d = R_x - R_y$.

3. Возводим в квадрат разности рангов d^2 и суммируем квадраты рангов.

4. Рассчитываем:

коэффициент ранговой корреляции по формуле:

$$r_s = 1 - \frac{6 \cdot \sum d^2}{n \cdot (n^2 - 1)} =$$

ошибку $S_{r_s} = \sqrt{\frac{1 - r_s^2}{n - 2}} =$

критерий значимости $t_{r_s} = \frac{r_s}{S_{r_s}} =$

Вывод:

Оценка результатов:

Компетенции	Владение теорией	Владение практикой	Балл
Владеет методами оценки связи (зависимости) между изучаемыми признаками	+	+	2
Владеет на удовлетворительно	+-	+-	1
Не владеет	-	-	0

Дата _____ Подпись преподавателя _____

Корреляционно-регрессионный анализ в программе Excel

Прямолинейная корреляция в программе Excel

Пример 1. Масса зерна ячменя (X, мг) и содержание жира в зерне (Y,%)

X – масса зерна, мг	11,0	19,9	15,9	16,3	10,2	21,4	15,8	21,6	12,3	17,3
Y – содержание жира, %	1,2	5,1	2,3	3,1	0,9	4,1	2,1	4,2	1,1	3,4

1. В активный лист программы Excel введем исходные данные вышеприведенного примера, расположив значения по столбцам (рис. 7.1).
2. Из **Пакета анализа** выберем инструмент **Корреляция**
3. В появившемся окне укажем входной интервал **A2:B12**
4. **Группирование** по столбцам
5. Укажем метки и выходной интервал и нажмем **ОК** (рис. 7.1).

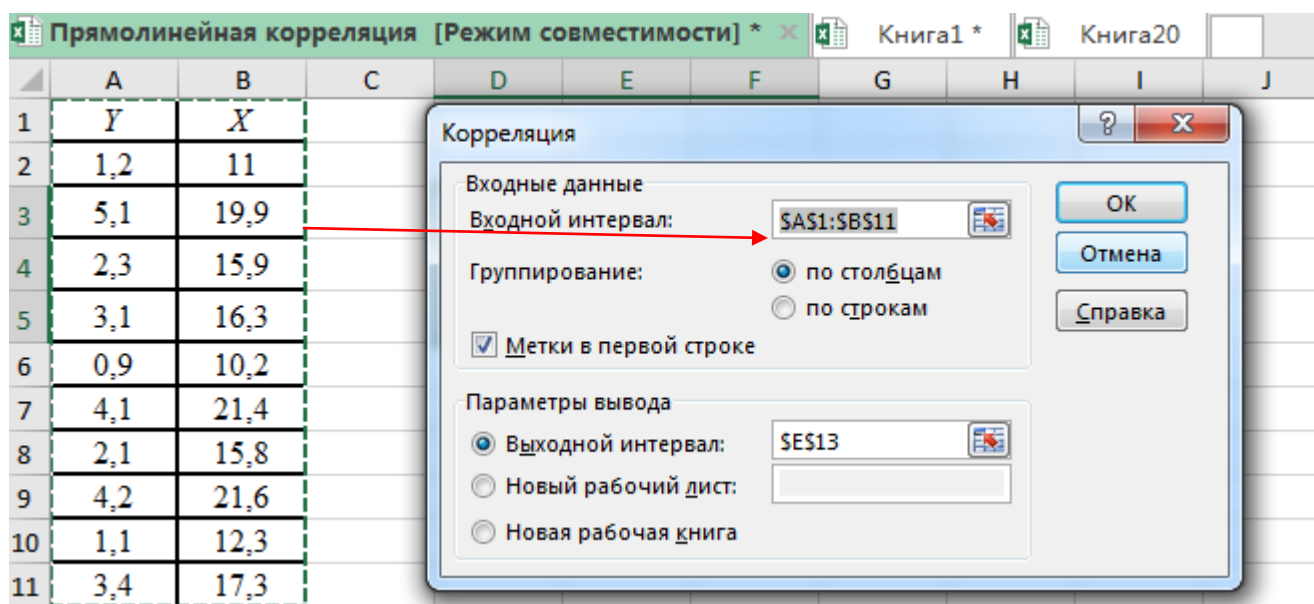


Рис. 7.1. Диалоговое окно *Корреляция*

После нажатия на клавишу **ОК** получаем коэффициент корреляции между массой зерна и содержанием жира в зерне ($r = 0,93$), который указывает на сильную (тесную) корреляционную зависимость между изучаемыми признаками (рис. 7.2).

	A	B	C	D	E	F
1	Y	X				
2	1,2	11				
3	5,1	19,9			Y	X
4	2,3	15,9		Y	1	
5	3,1	16,3		X	0,939	1
6	0,9	10,2				
7	4,1	21,4				
8	2,1	15,8				
9	4,2	21,6				
10	1,1	12,3				
11	3,4	17,3				

Рис. 7.2. Коэффициент корреляции между содержанием жира и массой зерна

Прямолинейная регрессия в программе Excel

Для расчета коэффициента регрессии, определения уравнения регрессии и построения теоретической линии в подменю **Анализ данных** выберем инструмент анализа **Регрессия** (рис. 7.3).

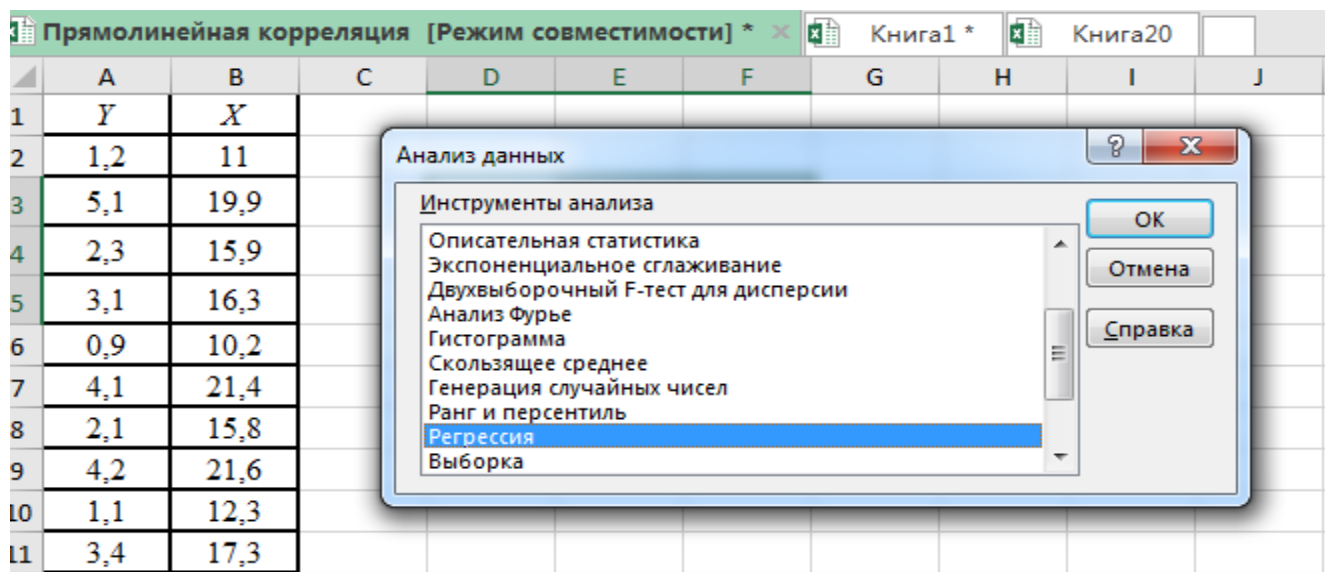


Рис. 7.3. Исходные данные и подменю *Анализ данных* – выбор регрессии

В появившемся окне настроек регрессии в поле **Входной интервал Y:** введем с помощью мышки диапазон ячеек, где расположены переменные данные зависимого признака (содержание жира) **A1:A11**, в поле **Входной интервал X:** введем с помощью мышки диапазон ячеек **B1:B12** – данные по массе зерна, галочкой отмечаем **Метки**. В поле **Уровень надежности** укажем доверительную вероят-

ность **95%**, отметим выходной интервал для размещения результатов расчета, в нашем случае – новый рабочий лист, выберем **График подбора** и нажмем на клавишу **ОК** (рис. 7.4).

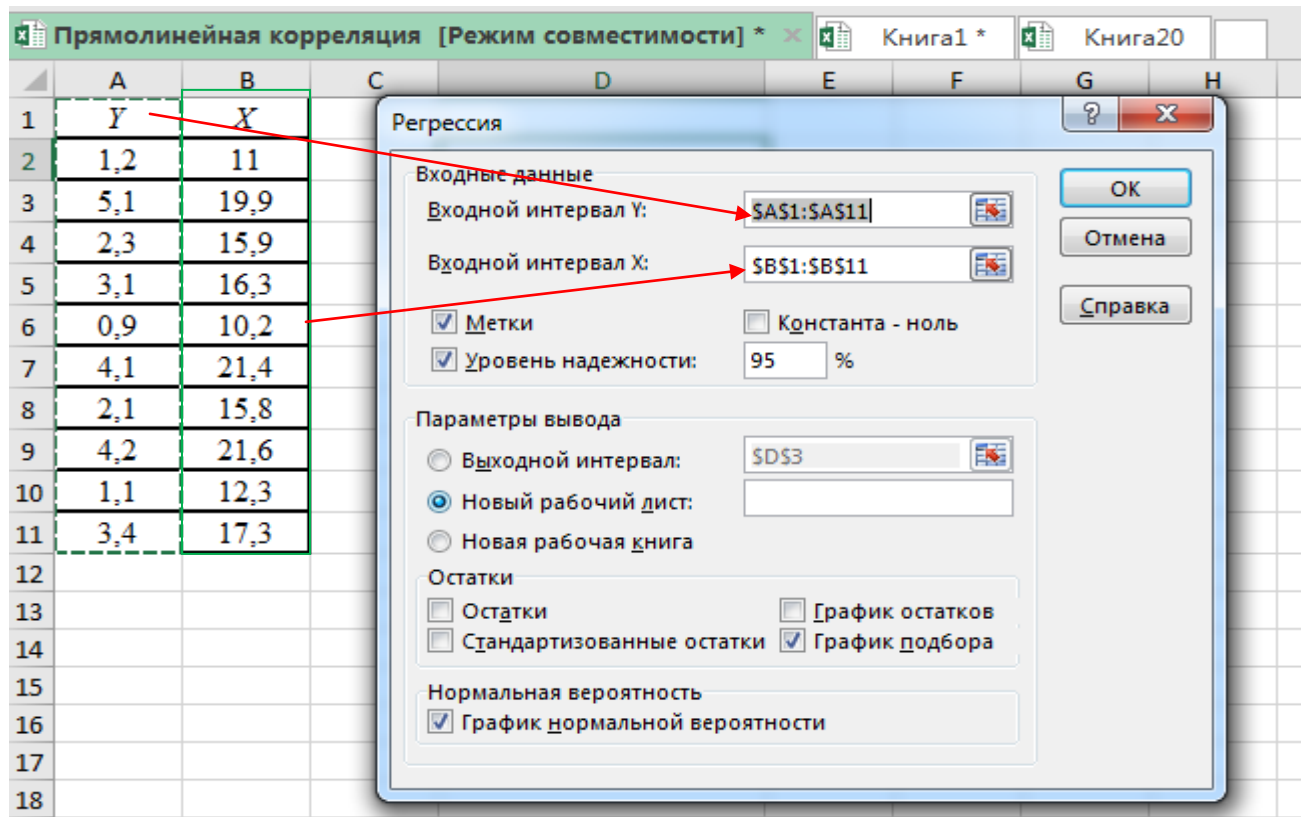


Рис. 7.4. Диалоговое окно *Регрессия*

После нажатия на клавишу **ОК** получаем на рабочем листе: **Вывод итогов, Вывод остатков, Дисперсионный анализ регрессии и График зависимости Y от X** (рис. 7.5).

Коэффициент корреляции ($r=0,94$) свидетельствует о тесной зависимости между изучаемыми признаками, коэффициент детерминации ($d_{yx} = r^2 = 0,88$) показывает, что изменение содержания жира на 88 % зависит от массы зерна.

Дисперсионный анализ регрессии дает возможность провести оценку качества модели уравнения регрессии. Данные таблицы показывают, что из общей суммы квадратов отклонений (SS), равной $19,165$ на регрессию (влияние массы зерна на содержание жира) приходится $16,903$ (88,3%) и $2,263$ на остаточную или случайную вариацию. F – критерий полученного уравнения регрессии значим на 5% уровне ($F_{\text{фак}} > F_{01}$). Вероятность отклонения нулевой гипотезы $p = 0,000056$ значительно меньше $0,01$, что говорит об общей значимости предлагаемой модели регрессии.

В нижней части таблицы указаны параметры для нахождения уравнения линейной регрессии: $b_{yx} = 0,336$ – коэффициент регрессии; $a = -2,685$ – свободный член. Уравнение регрессии будет иметь следующий вид: $Y = 0,34X - 2,68$

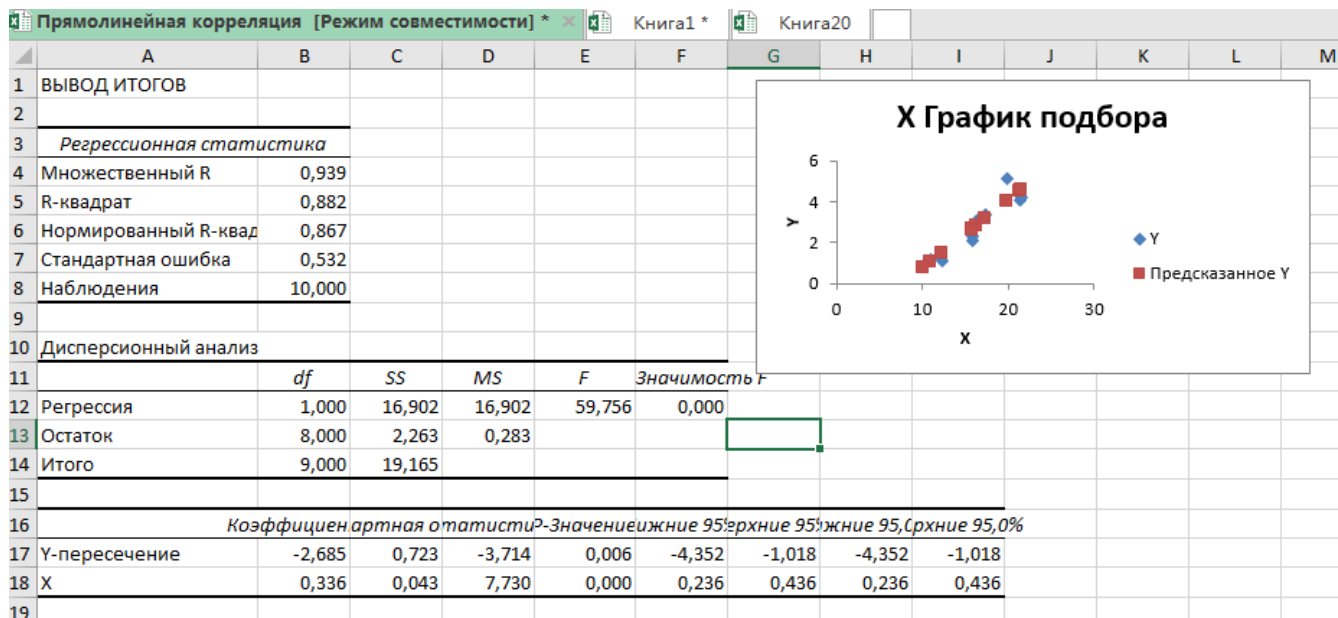


Рис. 7.5. Итоги регрессионного анализа

Корреляционное поле и теоретическая линия регрессии.

Так как график корреляции на рабочем листе выглядит неразборчиво, с помощью стандартной операции, используя левую и правую клавишу мышки, перенесем график на отдельный лист, растянем его вширь, в высоту и отформатируем его (рис 7.6.).

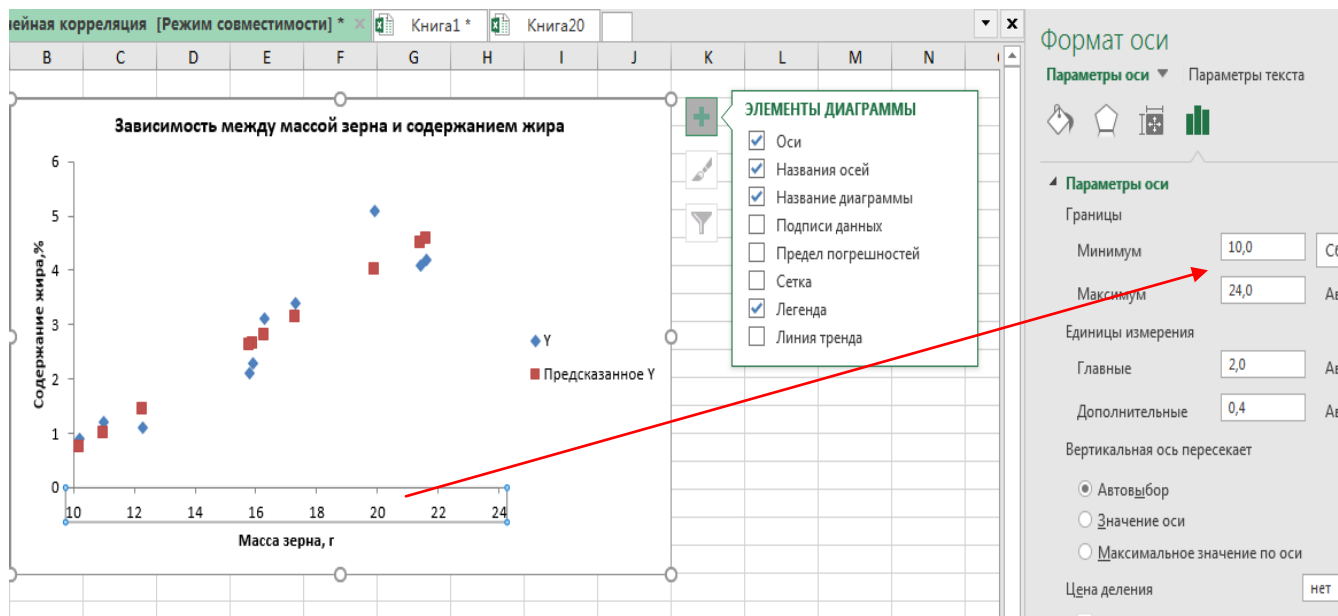


Рис. 7.6. Диалоговое окно для форматирования элементов графика

Для построения теоретической линии и нахождения уравнения регрессии подведем курсор мышки к данным графика (фактическим точкам жира), при нажатии на правую клавишу появляется контекстное меню, выберем **Добавить линию тренда** (рис. 7.7)

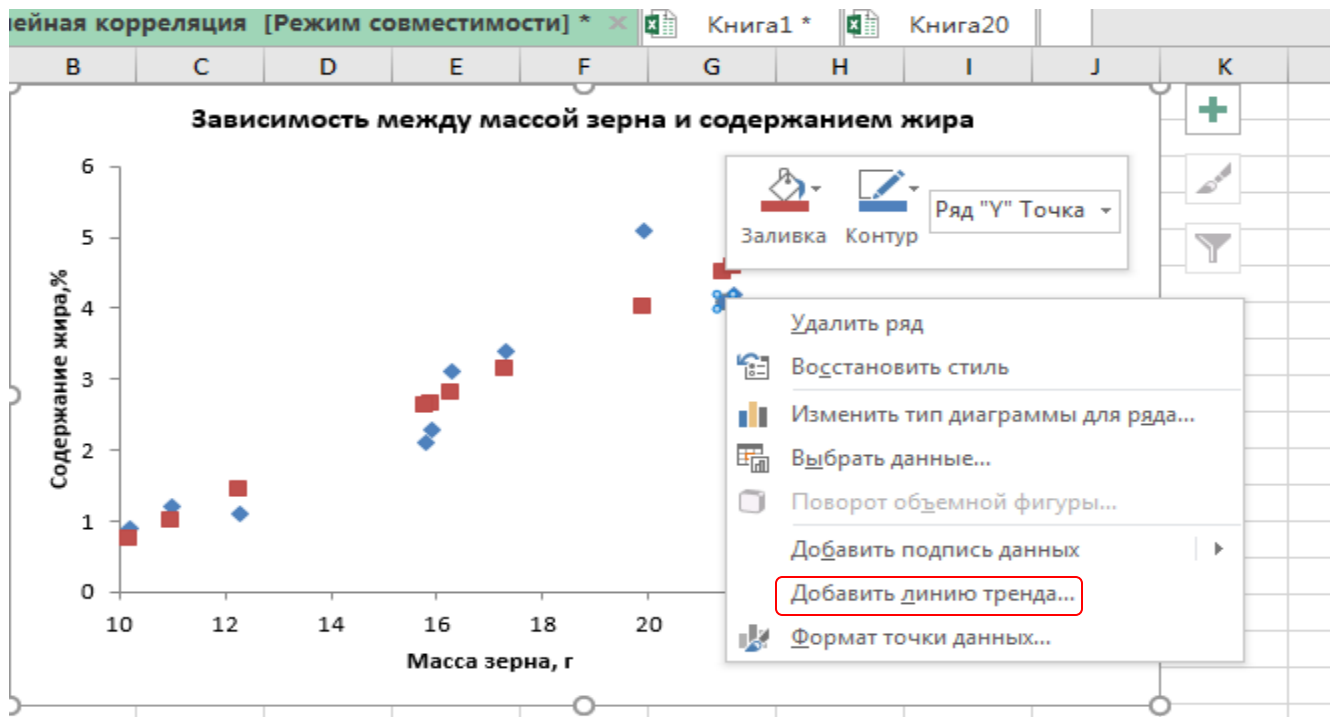


Рис. 7.7. Контекстное меню для работы с данными

После выбора **Добавить линию тренда** появляется всплывающее окно с параметрами линии тренда. Так как мы предполагаем, что наша зависимость носит прямолинейный характер, для аппроксимации и сглаживания эмпирической линии выберем **Линейная**, далее выберем **автоматическое сглаживание**, галочкой укажем **показать уравнение на диаграмме** и поместить **R²** (рис.7.8).

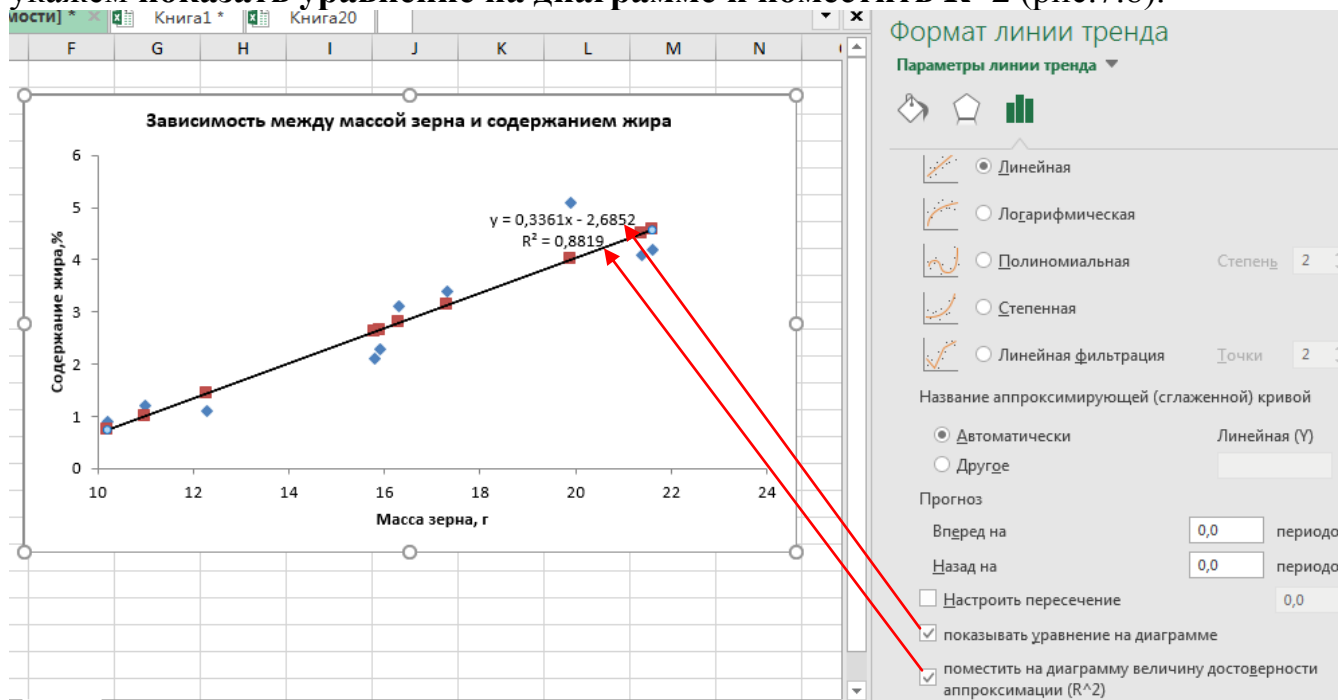


Рис. 7.8. Формат линии тренда

После выбора параметров линии тренда получаем окончательный график зависимости между содержанием жира и массой зерна с наименованием осей Y , X и расшифровкой легенды графика (рис. 7,9). На графике голубыми кубиками отмечено фактическое содержание жира, оранжевыми – предсказанное или теоретическое содержание жира и уравнение регрессии: $Y = 0,34X - 2,68$. Из данных таблицы на рис.7.5, так и из уравнения регрессии видно, что коэффициент регрессии составляет 0,34%. Данный коэффициент свидетельствует о том, что при увеличении массы зерна на 1 г, содержание жира увеличивается на 0,34%.

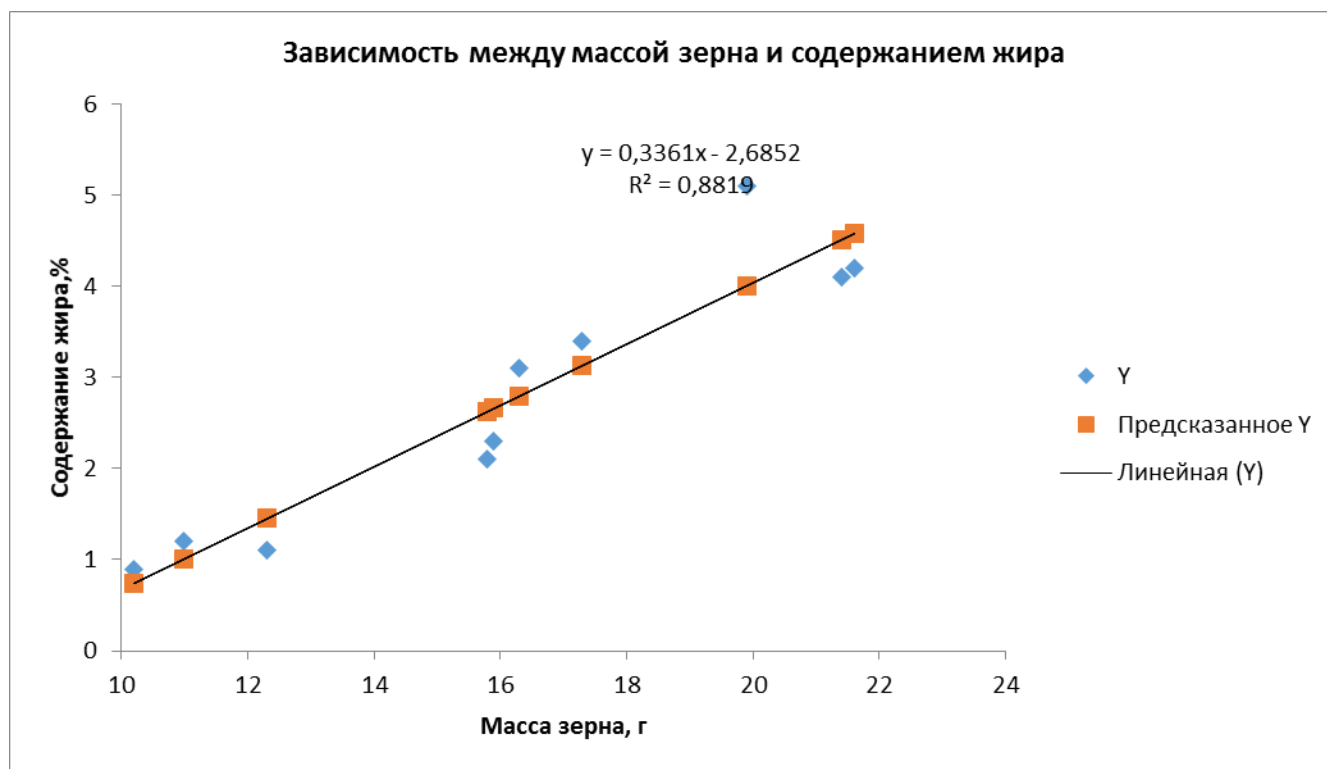


Рис. 7.9. График зависимости между содержанием жира и массой зерна

Криволинейная регрессия в программе Excel

В Пакете данных программы Excel можно рассчитать только параметры прямолинейной корреляции и регрессии, а инструментов для нахождения параметров нелинейной регрессии нет. Для оценки нелинейной зависимости важной задачей является подбор вида функции, которая бы наиболее точно описывала искомую зависимость. Эта задача в какой-то мере может быть решена инструментами **Диаграммы**.

Пример. Степень повреждения побегов черной смородины второго поколения листовой галлицей $Y(\%)$ в зависимости от количества эфирных железок на листе X (шт).

X	48	50	53	55	58	61	60	72	73	51
Y	69	82	69	50	78	37	54	38	62	64
X	59	83	107	122	128	129	137	197	176	131
Y	68	34	32	21	25	20	15	14	13	17

1. В активный лист программы Excel введем исходные данные вышеприведенного задания, расположив данные по столбцам (рис. 7.11).

2. Выделим левой кнопкой мыши диапазон ячеек **A1:B21**, далее правой кнопкой нажимаем на вкладку «**Вставка**», далее **Диаграмма**. В появившемся окне выберем тип **Точечная**. Результат обработки появится в виде диаграммы.

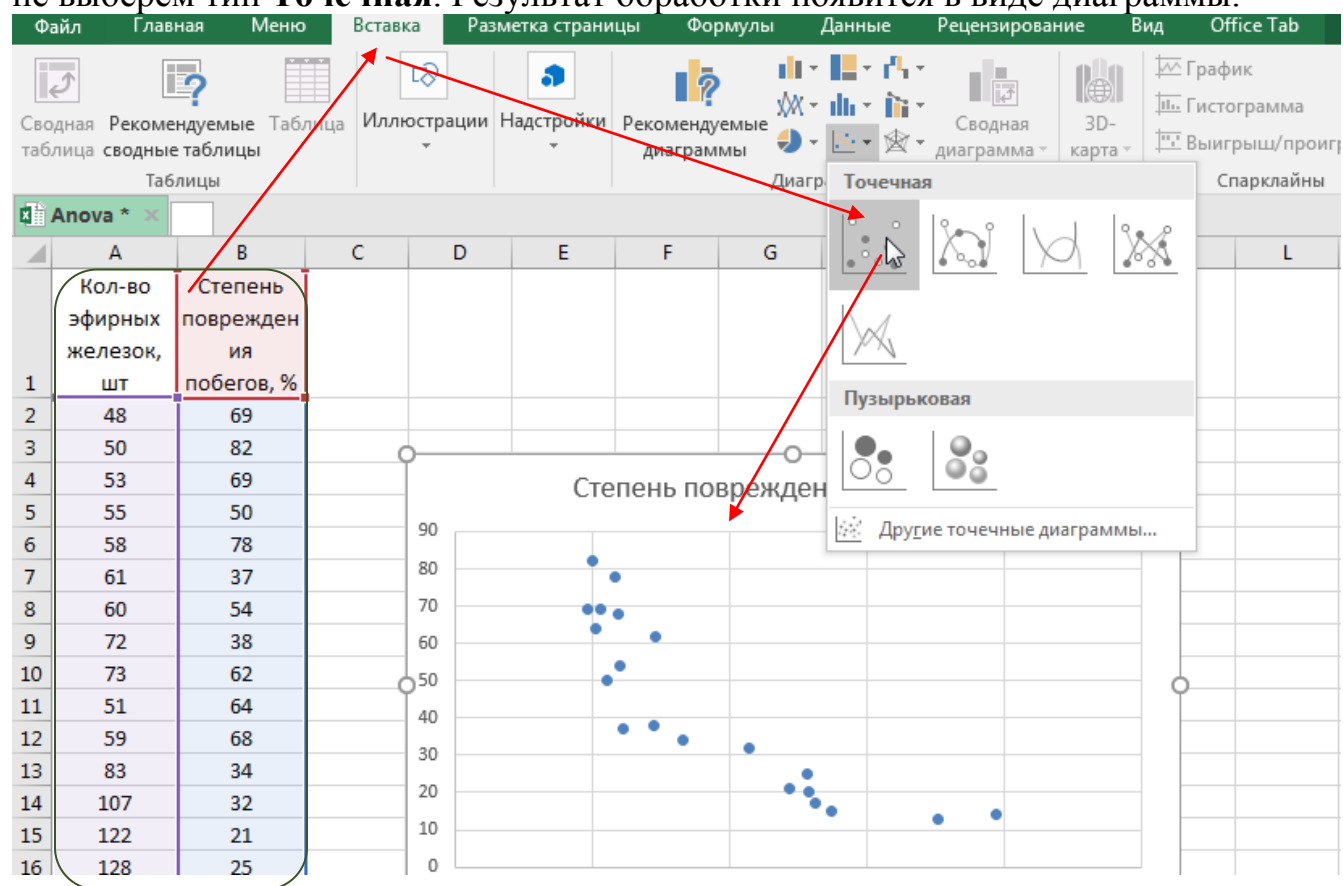


Рис. 7.11. Таблица исходных данных и выбор точечного типа диаграммы

3. Так как график на рабочем листе выглядит неразборчиво, с помощью стандартной операции, используя левую и правую клавишу мышки, перенесем график на отдельный лист, растянем его вширь, в высоту и проведем форматирование диаграммы: изменение масштаба вертикальной оси, наименование графика, наименование горизонтальной и вертикальной осей и т.д.

4. Форматирование рабочей области диаграммы и ее элементов проводится в контекстном меню с помощью левой или правой клавиш мыши.

5. Так как программа автоматически устанавливает минимальное значение вертикальной и горизонтальной осей графика равное нулю, эмпирические точки часто располагаются, как в нашем случае, только в верхней или правой части графика (рис. 7.11). Из-за этого искажается визуальная оценка графика, поэтому рекомендуется форматирование осей диаграммы. Для этого левой клавишей мышки выделим горизонтальную ось, в правой части появляется панель **Формат оси**, в окошке границ минимума заменим значение 0 на 40 (в нашем примере минимальное значение количества эфирных железок равно 48). Автоматически вертикальная ось растягивается в интервале от 40 до 250 (рис. 7.12).

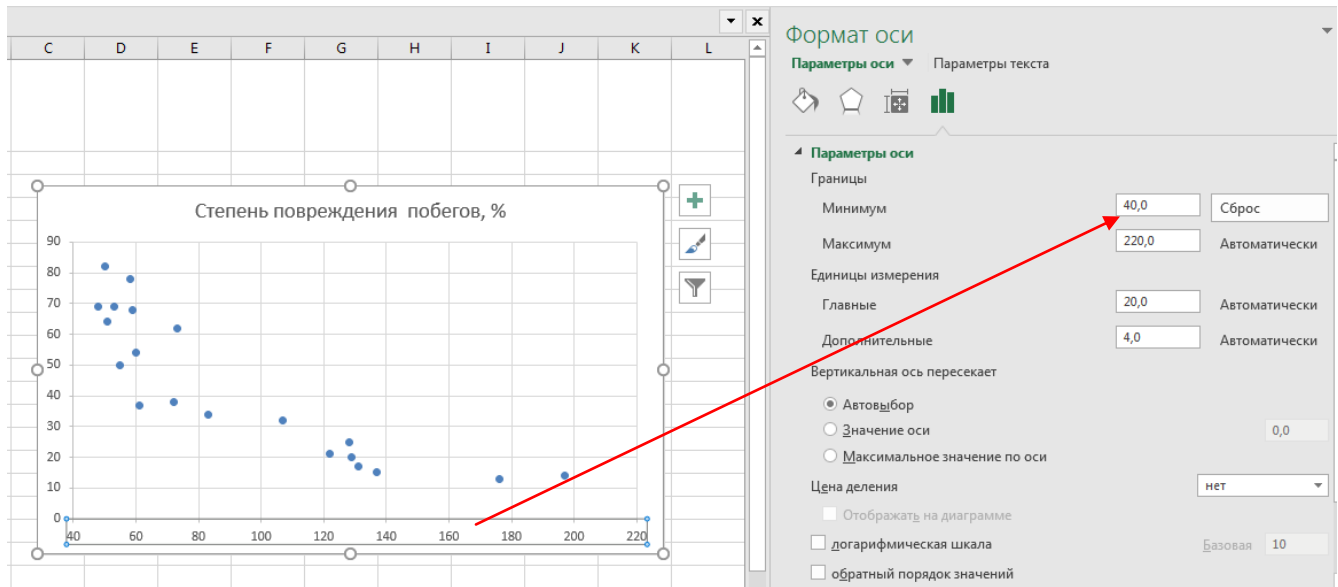


Рис. 7.12. Форматирование вертикальной оси диаграммы

6. Для построения теоретической линии и нахождения уравнения регрессии подведем курсор мыши к любой точке диаграммы, при нажатии на правую клавишу появляется контекстное меню для форматирования данных диаграммы, выберем **Добавить линию тренда** (рис. 7.13).

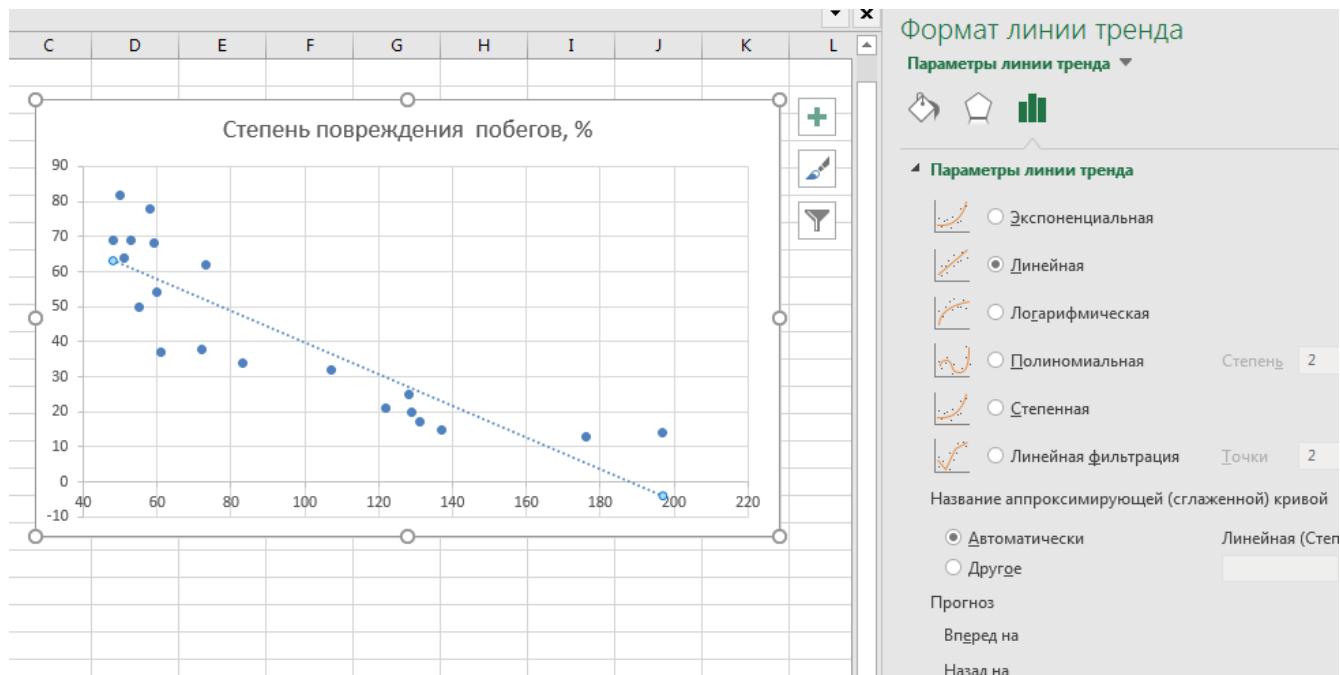


Рис. 7.13. Контекстное меню для форматирования данных диаграммы

7. Появляется контекстное меню для выбора параметров линии тренда. В меню представлены 6 типов аппроксимации: По умолчанию программа выбирает линейную модель. Основными типами нелинейных регрессий являются: экспоненциальная, логарифмическая, полиномиальная и степенная.

8. По умолчанию программа выбирает линейную модель, галочкой укажем показать уравнение на диаграмме и поместить на диаграмму величину достоверности (R^2), на диаграмме автоматически появится уравнение прямой $Y = -0,45X + 84,728$ и значение коэффициента детерминации $R^2 = 0,7478$ (рис.7.14).

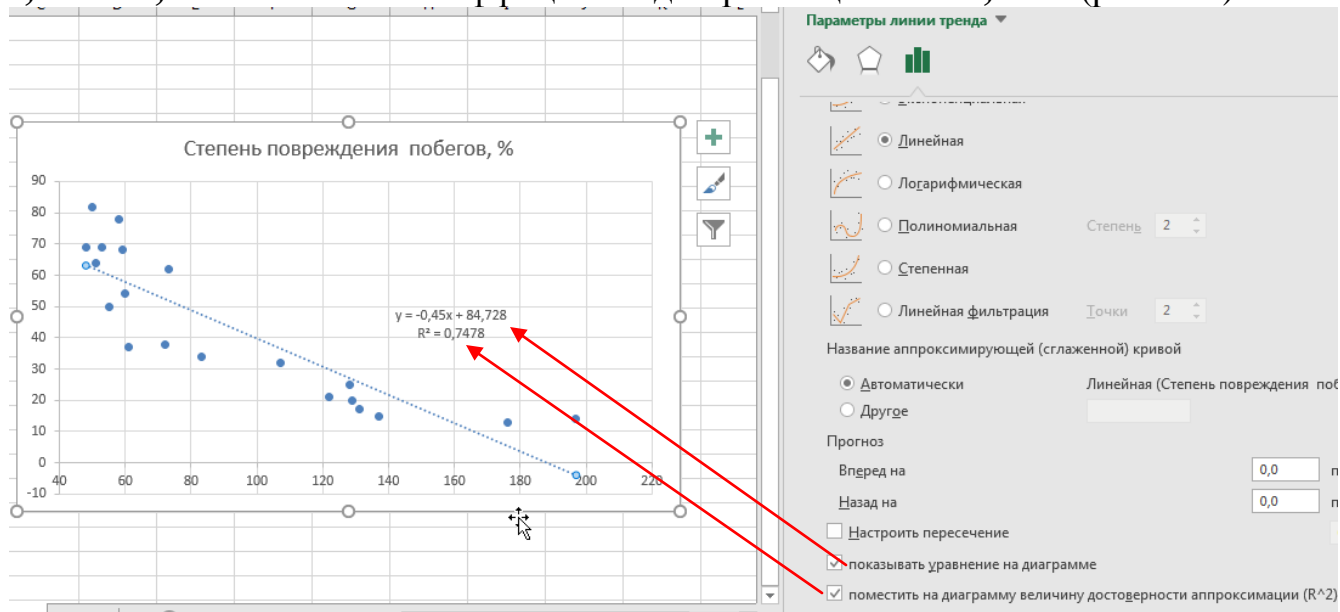
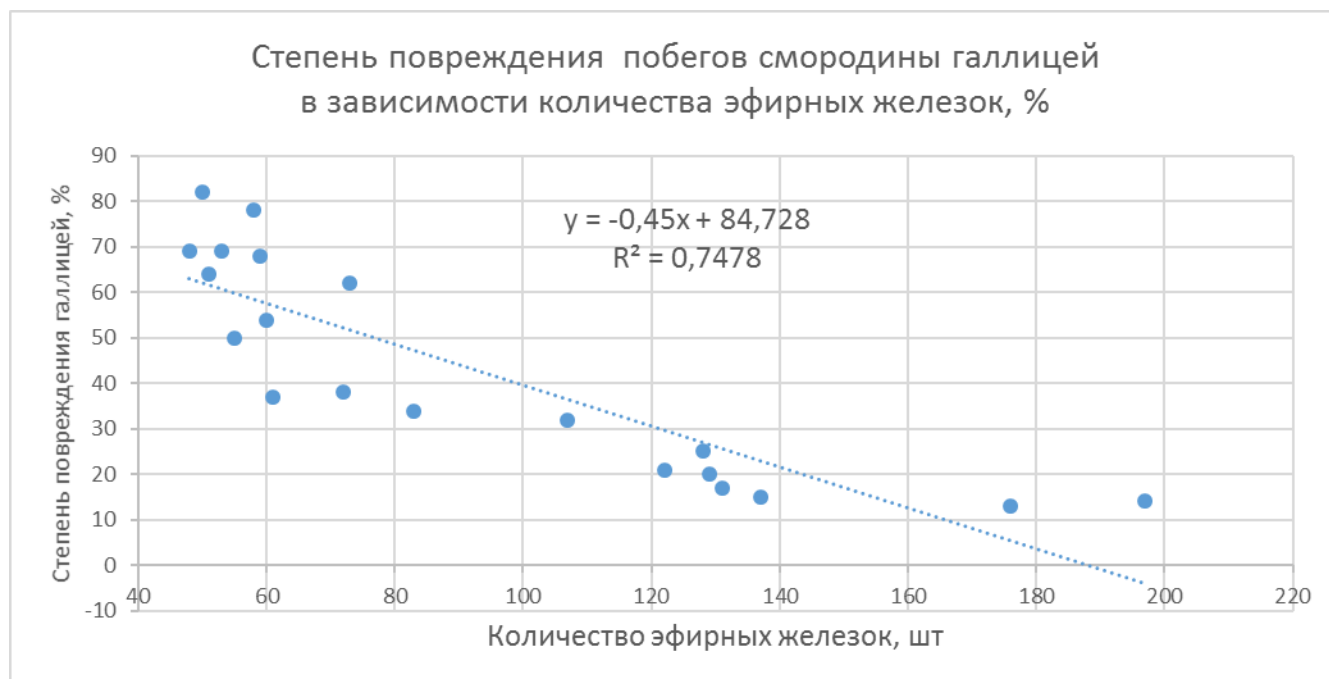


Рис.7.14. Контекстное меню для форматирования параметров линии тренда

9. С помощью панели ЭЛЕМЕНТЫ ДИАГРАММЫ укажем наименование графика, наименование осей, как показано на рис. 7.15.



7.15. Отредактированная диаграмма

10. Визуальная оценка данных корреляционного поля показывает, что вероятнее всего зависимость повреждения побегов смородины галлицей в зависимости от количества эфирных железок носит нелинейный характер. Поэтому проведем аппроксимацию по всем типам представленных в программе Excel криволинейных регрессий. Одним из основных показателей аппроксимации является R^2 или коэффициент детерминации, по которому можно судить о качестве модели. Чем больше значение R-квадрат, тем лучше аппроксимируется выбранная модель, тем лучше качество модели.

11. Для изменения типа моделей криволинейной регрессии подводим курсор мыши к линии прямолинейной регрессии (рис. 7.15), далее в меню параметров линии трендов последовательно выбираем (рис. 7.14) экспоненциальную, логарифмическую, степенную и показательные регрессии.

12. Для выбора параметров уравнения регрессии и лучшей модели нелинейной зависимости проанализируем 4 диаграммы итогов регрессии, представленных ниже (рис. 7.16;7.17;7.18;7.19). Лучшая аппроксимация достигается при более высоких значениях коэффициента детерминации (R^2).

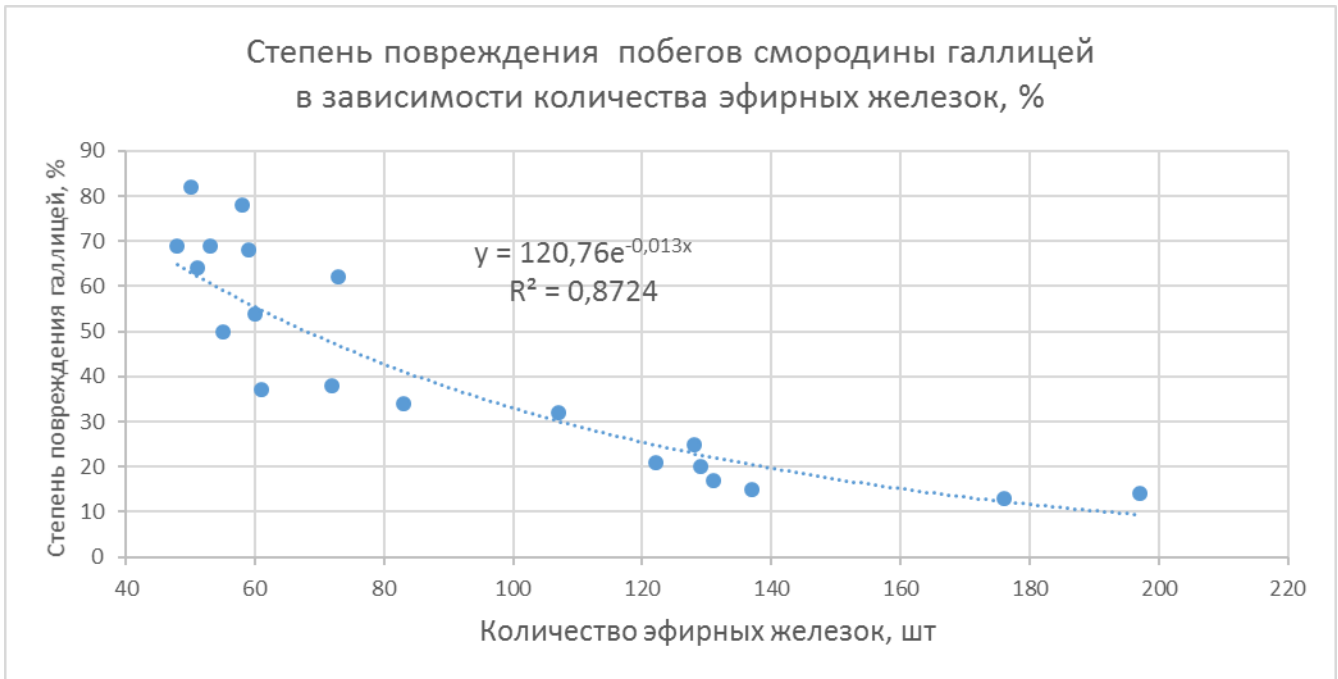


Рис. 7.16. Экспоненциальная модель

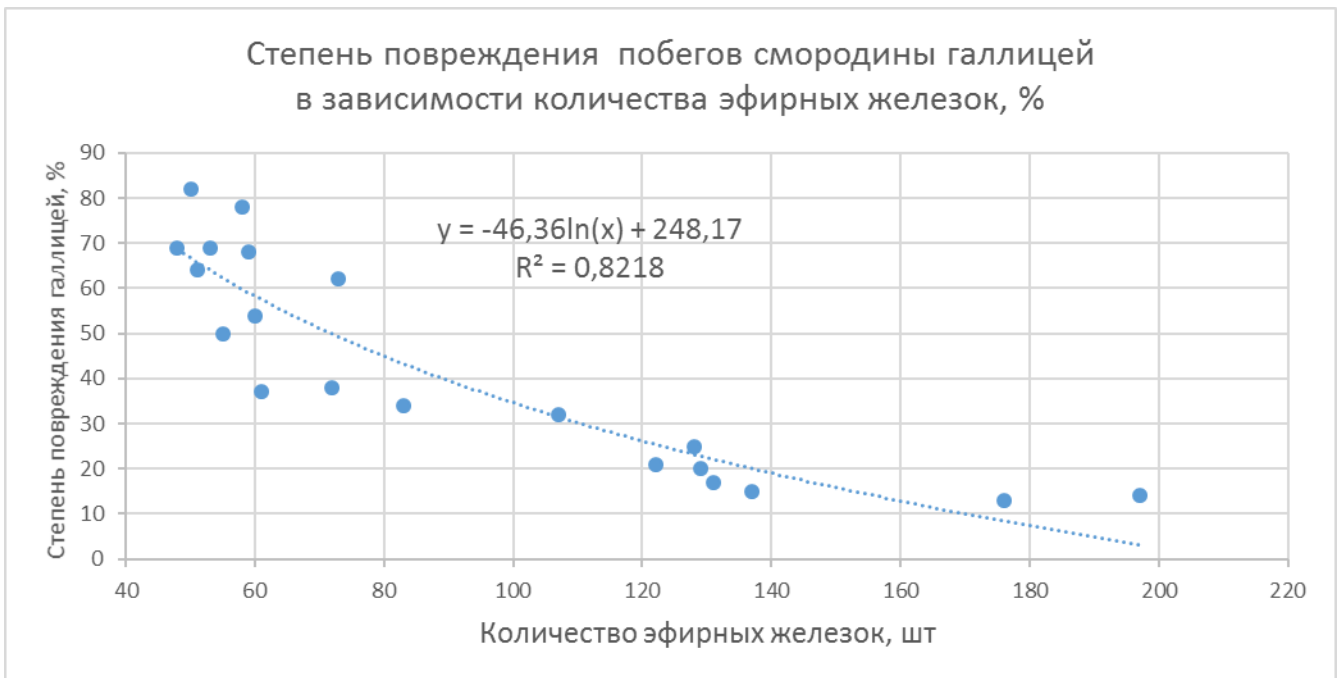


Рис. 7.17. Логарифмическая модель

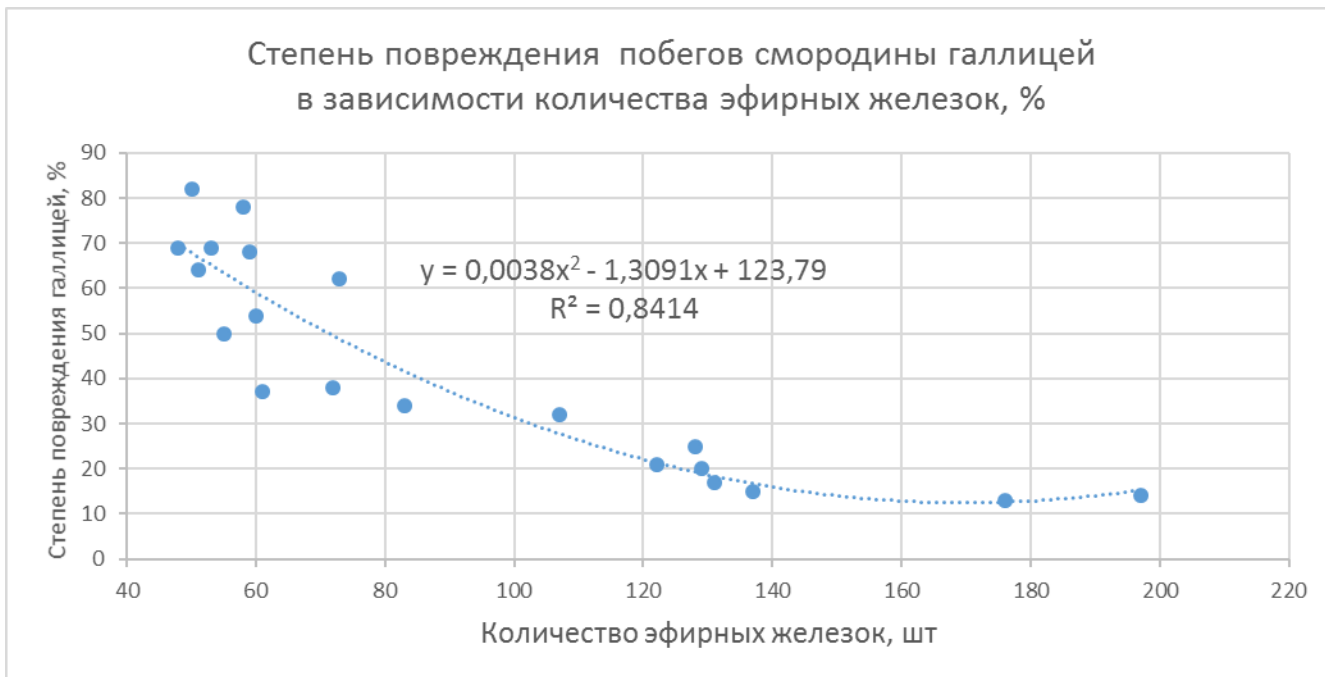


Рис. 7.18. Показательная модель

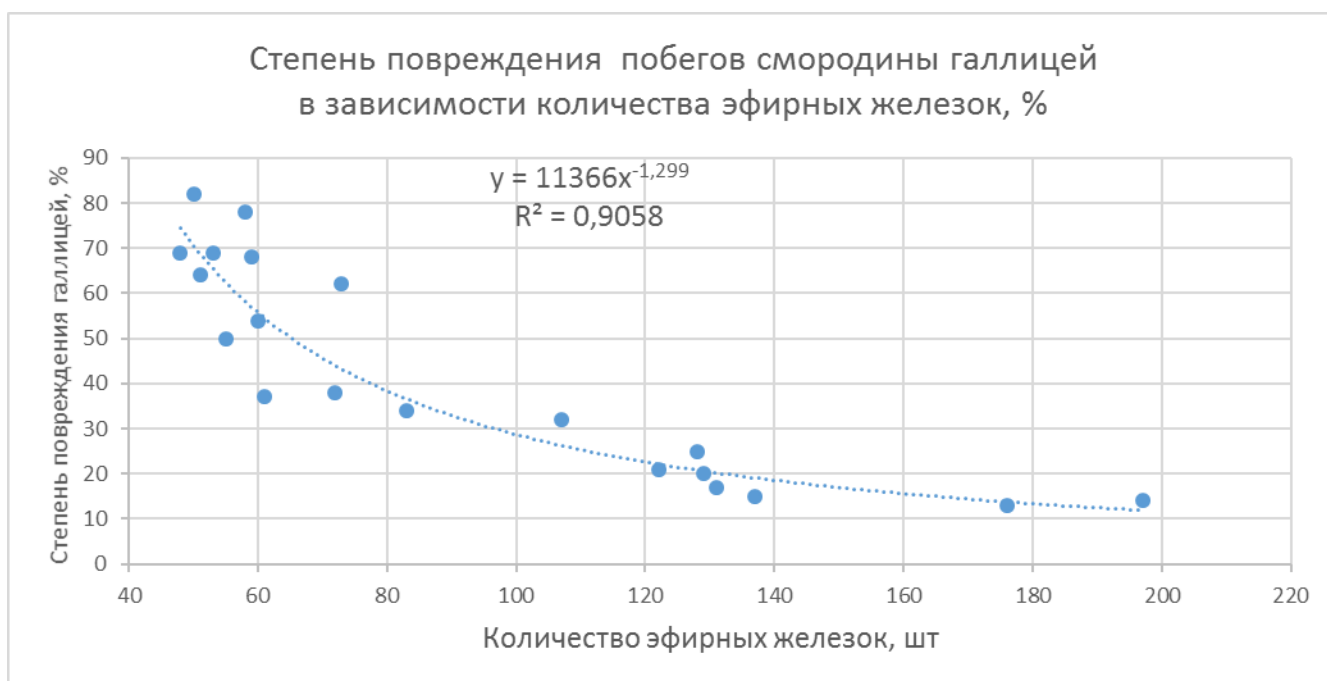


Рис. 7.19. Степенная модель

Наиболее точно в нашем примере зависимость степени повреждения побегов черной смородины листовой галлицей от количества эфирных железок на листе описывается степенной моделью $Y = 11366x^{-1,299}$; $R^2 = 0,91$ (рис.7.19).

Работа 8. Дисперсионный анализ данных экспериментов с полной рандомизацией вариантов

Теоретическая часть:

При изучении причинно-следственных отношений между явлениями особенно ценным оказался метод *дисперсионного анализа*. Этот метод основан на разложении общей дисперсии статистического комплекса на составляющие компоненты (отсюда и название метода), сравнивая, которые друг с другом посредством критерия Фишера (F - критерия), можно определить долю общей вариации изучаемого (результативного) признака, обусловленную действием на него, как регулируемых, так и не регулируемых в опыте факторов.

Сущностью дисперсионного анализа является одновременное разложение суммы квадратов и числа степеней свободы на составляющие компоненты, которые соответствуют структуре эксперимента и оценка действия и взаимодействия изучаемых вариантов по F - критерию. В зависимости от условий проведения опытов применяют различные схемы (модели) дисперсионного анализа, в которых записывают на какие конкретно суммы квадратов и степени свободы расчленяют общее варьирование.

Лабораторные и вегетационные эксперименты чаще всего представляют собой статистические комплексы, в которых независимость сопоставимых вариантов достигается регулярным перемещением сосудов на столах или вагонетках.

Независимости можно добиться и в полевых условиях, если варианты по делянкам полевого опыта размещать методом полной рандомизации. Следовательно, в лабораторных, вегетационных экспериментах и в полевых опытах с полной рандомизацией вариантов нет территориально организованных повторений. В таких случаях дисперсионный анализ данных указанных экспериментов необходимо вести как для несопряженных (независимых) выборок.

Общее варьирование изучаемого признака будет состоять из варьирования между выборками (вариантами) и варьирования внутри выборок. Вариация между выборками (вариантами) представляет ту часть общей изменчивости, которая обусловлена действием изучаемых факторов, а вариация внутри выборок характеризует действие случайных факторов, т.е. ошибку эксперимента.

Оценка значимости действия изучаемых факторов проводится по *критерию Фишера* – F , который представляет собой отношение дисперсии (среднего квадрата) вариантов к дисперсии ошибки.

Если $F_{\text{факт}} < F_{\text{табл.}}$, то нулевая гипотеза (предположение: все средние значения по вариантам являются оценками одной генеральной средней и между нет существенных различий) принимается и на этом расчеты заканчиваются, если $F_{\text{факт}} \geq F_{\text{табл.}}$, то нулевая гипотеза отвергается и необходимо дополнительно оценить существенность частных различий по величине HSP_{05} или HSP_{01} и определить между какими средними имеются значимые различия.

Контрольные вопросы:

1. Сущность дисперсионного анализа.	2. Модель дисперсионного анализа данных однофакторного опыта с полной рандомизацией вариантов.
2. Проверка нулевой гипотезы при дисперсионном анализе.	4. Особенности дисперсионного анализа опыта с разной повторностью вариантов (выпавшими датами)

Общая постановка задачи:

- изучить сущность дисперсионного анализа, этапы и условия его применения;
- освоить практические навыки проведения дисперсионного анализа урожайных данных вегетационных опытов;
- сделать статистический и биологический вывод.

Список индивидуальных заданий

Каждый пример работы 8 включает 32 задания, которые получены путем различных сочетаний 4-х из 7-ми повторностей опыта.

Запишите данные своего задания по повторностям в исходную таблицу и проведите расчеты.

Шифр задания к работе 8

№	Повторности	№	Повторности	№	Повторности	№	Повторности
1	1 2 3 4	9	2 3 5 7	17	1 3 5 7	25	1 3 4 7
2	2 3 4 5	10	1 2 4 7	18	2 5 6 7	26	1 2 6 7
3	3 4 5 6	11	2 4 5 6	19	1 2 5 6	27	2 3 6 7
4	1 2 3 5	12	1 4 5 6	20	1 3 4 5	28	2 3 4 7
5	4 5 6 7	13	3 5 6 7	21	1 4 5 7	29	3 4 5 7
6	1 2 3 6	14	1 3 6 7	22	1 4 6 7	30	2 4 5 7
7	2 3 5 6	15	1 3 5 6	23	1 5 6 7	31	1 3 4 6
8	1 2 3 7	16	2 4 6 7	24	1 4 5 7	32	2 4 7

Примечание: При выполнении задания с выпавшими датами (разной повторностью по вариантам) исключить значения данных второго варианта (в 4-ой повторности) и значений третьего варианта (в 3-ей и 4-ой повторности).

Пример 1. В опыте по биотестированию наночастиц изучали влияние концентрации наночастиц серебра NPAg (моль/л) на длину корешков проростков томата, (мм). Экспозиция 30 мин.

Таблица 26

Концентрация наночастиц NPAg (моль/л)	Повторности						
	1	2	3	4	5	6	7
1. Контроль, st	35,4	37,1	36,6	35,8	36,0	37,1	36,6
2. NPAg, $16 \cdot 10^{-7}$	35,6	38,8	37,8	36,5	37,6	38,2	37,1
3. NPAg, $8 \cdot 10^{-6}$	39,6	40,4	38,3	39,5	41,1	40,7	39,5
4. NPAg, $4 \cdot 10^{-5}$	41,6	39,9	39,3	45,6	40,8	42,7	43,1
5. NPAg, $2 \cdot 10^{-4}$	46,2	45,4	43,8	49,9	50,0	47,2	45,5

Пример 2. Влияние сред разного гормонального состава на содержание суммы растворимых фенольных соединений (СРФС) в листьях клюквы крупноплодной, мг/г сырой массы

Таблица 27

Варианты опыта	Повторности						
	1	2	3	4	5	6	7
1. Без гормонов	50	53	55	51	57	59	55
2. ИУК/Кин	32	31	34	35	37	33	34
3. ИУК/БАП	33	35	32	34	35	32	29
4. ИМК/Кин	21	19	25	23	25	24	22
5. ИМК/БАП	25	24	28	27	26	25	27

Ауксины: ИУК – индолилуксусная кислота, ИМК – индолилмасляная кислота;
Цитокинины: Кин – кинетин, БАП – бензиламинопурин

А. Одинаковая повторность вариантов (полный набор дат)

Пример № ____ задание _____

Таблица 28

Исходная таблица

Варианты опыта	Повторности				Суммы V	Средние \bar{x}_v
	1	2	3	4		
1.						
2.						
3.						
4.						
5.						

$$\sum X = \sum V =$$

Таблица 29

Вспомогательная таблица для вычисления сумм квадратов отклонений

Варианты опыта	X^2				V^2
	1	2	3	4	
1.					
2.					
3.					
4.					
5.					

$$\sum X^2 =$$

Общее число наблюдений $N = v n =$

Поправка $C = \frac{(\sum X)^2}{N} =$

Суммы квадратов отклонений:

общая $CKO = \sum X^2 - C =$

вариантов $CKV = \frac{\sum V^2}{n} - C =$

остаток $SKE = CKO - CKV =$

Таблица дисперсионного анализа

Источник вариации	Суммы квадратов, $СК$	Степени свободы, df	Средний квадрат, S^2	Критерий F	
				факт.	05
Общая			–	–	–
Вариантов					
Остаточная (ошибка)				–	–

$$s_x = \sqrt{\frac{S_e^2}{n}} =$$

Ошибки разности и HCP :

1) для сравнения со стандартом

$$S_d^I = \sqrt{\frac{2S_e^2}{n}} =$$

$$HCP_{05}^I = t_{05} \cdot S_d^I =$$

$$HCP_{01}^I = t_{01} \cdot S_d^I =$$

2) для сравнения с общей средней

$$S_d^{II} = \sqrt{S_e^2 \cdot \frac{(v-1)}{n \cdot v}} =$$

$$HCP_{05}^{II} = t_{05} \cdot S_d^{II} =$$

$$HCP_{01}^{II} = t_{01} \cdot S_d^{II} =$$

Таблица 31

Итоговая таблица сравнения средних по опыту _____

Варианты опыта	Средняя, \bar{x}_v	Отклонения от	
		стандарта, $d = \bar{x}_{on} - \bar{x}_{st}$	общей средней, $d = \bar{x}_{on} - \bar{x}_o$
1.			
2.			
3.			
4.			
5.			

Примечание: существенные отклонения пометить звездочками соответствующего уровня значимости.

Вывод:

В. Разная повторность вариантов или при наличии выпавших дат

Пример № _____ задание № _____

Таблица 32

Исходная таблица

Варианты опыта	Повторности, X				Суммы V	Средние \bar{x}_v
	1	2	3	4		
1.						
2.				–		
3.			–	–		
4.						
5.						

$$\sum X = \sum V =$$

Общее число наблюдений $N = v \cdot n - k =$

$$\text{Поправка } C = \frac{(\sum X)^2}{N} =$$

Суммы квадратов отклонений:

$$\text{общая } CKO = \sum X^2 - C =$$

$$\text{вариантов } CKV = \left(\frac{V_1^2}{n_1} + \frac{V_2^2}{n_2} + \dots + \frac{V_k^2}{n_k} \right) - C =$$

$$\text{остаток } CKE = CKO - CKV =$$

Таблица 33

Таблица дисперсионного анализа

Источник вариации	Суммы квадратов, СК	Степени свободы, df	Средний квадрат, S ²	Критерий F	
				факт.	05
Общая			–	–	–
Вариантов					
Остаточная (ошибка)				–	–

а) ошибка разности и HCP_{05} для сравнения вариантов с полным набором дат:

$$S_d^1 = \sqrt{\frac{2S_e^2}{n}} =$$

$$HCP_{05}^1 = t_{05} \cdot S_d^1 =$$

б) то же для сравнения вариантов с разной повторностью:

$$S_d^{II} = \sqrt{S_e^2 \frac{n_1 + n_2}{n_1 \cdot n_2}} =$$

$$HCP_{05}^{II} = t_{05} \cdot S_d^{II} =$$

$$S_d^{III} = \sqrt{S_e^2 \frac{n_1 + n_3}{n_1 \cdot n_3}} =$$

$$HCP_{05}^{III} = t_{05} \cdot S_d^{III} =$$

$$S_d^{IV} = \sqrt{S_e^2 \frac{n_2 + n_3}{n_2 \cdot n_3}} =$$

$$HCP_{05}^{IV} = t_{05} \cdot S_d^{IV} =$$

Таблица 34

Итоговая таблица сравнения средних по опыту _____

Варианты опыта	Повторность	Средние, \bar{x}_v	$d = \bar{x}_{it} - \bar{x}_{st}$	HCP_{05}
1.				
2.				
3.				
4.				
5.				

Вывод:

Оценка результатов:

Компетенции	Владение теорией	Владение практикой	Балл
Владеет навыками оценки результатов опыта на основе дисперсионного анализа	+	+	2
Владеет на удовлетворительно	+-	+-	1
Не владеет	-	-	0

Дата _____ Подпись преподавателя _____

Дисперсионный анализ данных опытов с полной рандомизацией вариантов в программе Excel

1. В активный лист программы Excel введем исходные данные примера по изучению влияния сред на содержание суммы растворимых фенольных соединений (СРФС) в листьях клюквы крупноплодной (рис. 8.1)

2. Из Пакета анализа выберем инструмент **Однофакторный дисперсионный анализ** (рис. 8.1).

Пример Влияние сред разного гормонального состава на содержание суммы растворимых фенольных соединений (СРФС) в листьях клюквы крупноплодной, мг/г сырой массы:

Варианты опыта	Повторность			
	1	2	3	4
1. Без гормонов	51	54	53	55
2. ИУК/Кин	33	32	35	33
3. ИУК/БАП	34	36	33	35
4. ИМК/Кин	20	19	23	22
5. ИМК/БАП	26	24	27	25

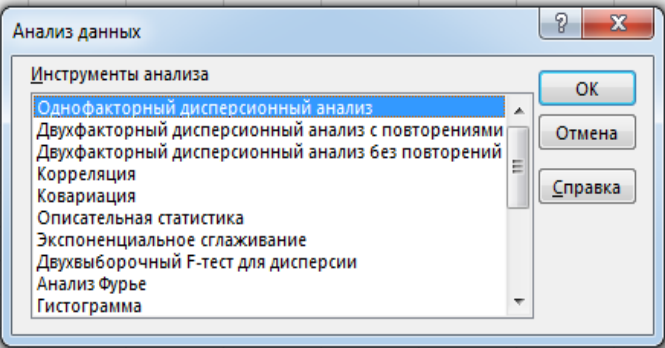


Рис. 8.1. Исходные данные и выбор инструмента для проведения дисперсионного анализа

3. В появившемся окне укажем входной интервал **A3:E7**. Входной интервал должен включать только диапазон, состоящий из перечня вариантов и цифровых данных по этим вариантам (рис.8.2).

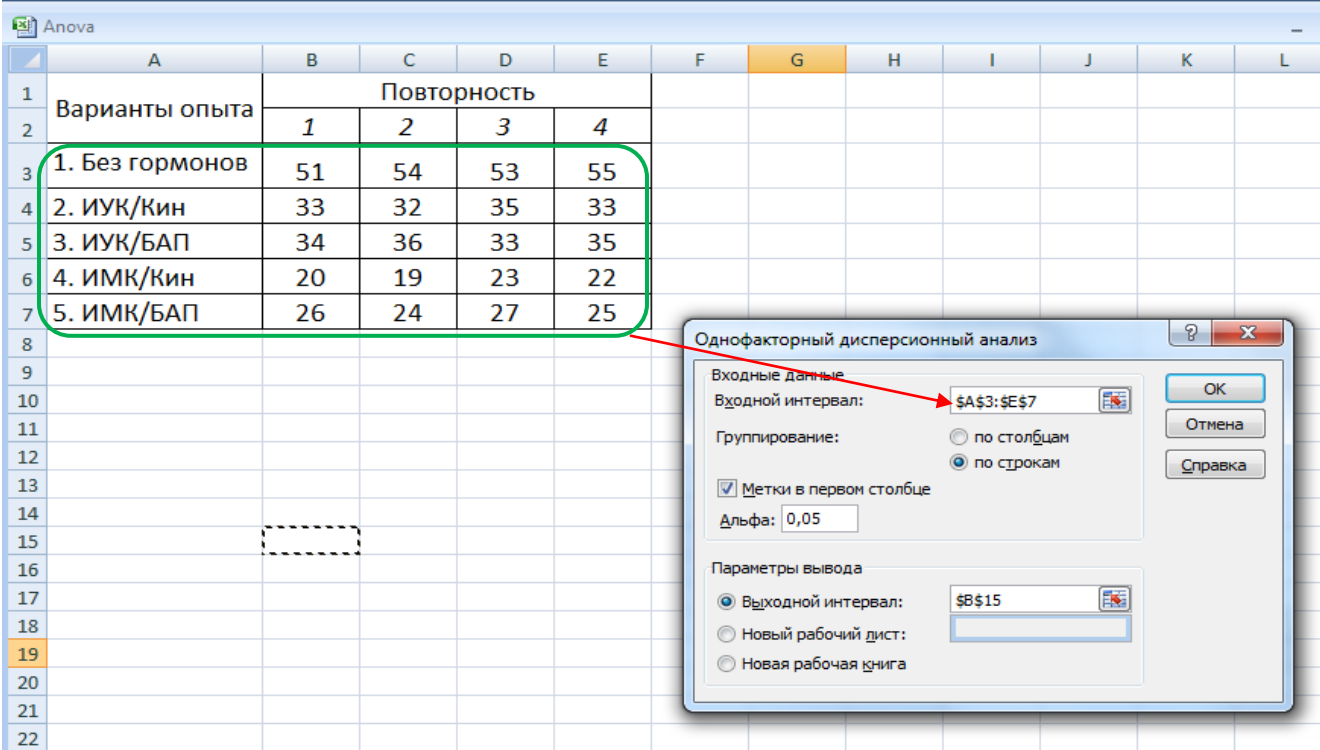


Рис. 8.2. Исходные данные и диалоговое окно **Однофакторный дисперсионный анализ**

4. Группирование – по строкам (рис.8.2).
5. Укажем метки в первом столбце (рис.8.2). Это необходимо для того, чтобы в выходных таблицах автоматически указывались наименования вариантов.
6. Альфа – выбор уровня значимости **0,05** или **0,01**
7. Выбираем выходной интервал для размещения результатов дисперсионного анализа: на данном листе или новом листе и нажимаем **ОК** (рис. 8.2)
8. Получаем итоговую таблицу дисперсионного анализа «Однофакторный дисперсионный анализ» (рис.8. 3)

Аноva										
	A	B	C	D	E	F	G	H	I	
1	Варианты опыта	Повторность								
2		1	2	3	4					
3	1. Без гормонов	51	54	53	55					
4	2. ИУК/Кин	33	32	35	33					
5	3. ИУК/БАП	34	36	33	35					
6	4. ИМК/Кин	20	19	23	22					
7	5. ИМК/БАП	26	24	27	25					
8										
9										
10	Однофакторный дисперсионный анализ									
11										
12	ИТОГИ									
13	<i>Группы</i>	<i>Счет</i>	<i>Сумма</i>	<i>Среднее</i>	<i>Дисперсия</i>					
14	1. Без гормонов	4	213	53,25	2,916667					
15	2. ИУК/Кин	4	133	33,25	1,583333					
16	3. ИУК/БАП	4	138	34,5	1,666667					
17	4. ИМК/Кин	4	84	21	3,333333					
18	5. ИМК/БАП	4	102	25,5	1,666667					
19										
20										
21	Дисперсионный анализ									
22	<i>Источник вариации</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-Значен</i>	<i>F критическое</i>			
23	Между группами	2445,5	4	611,375	273,75	8,03E-14	3,055568			
24	Внутри групп	33,5	15	2,233333						
25										
26	Итого	2479	19							
27										

Рис. 8.3. Итоговая таблица дисперсионного анализа

9. В верхней части итоговой таблицы Excel под **Группами** подразумеваются «Варианты», **Счет** – это повторность каждого варианта (n =4).

10. В нижней части таблицы термины и формулы подразумевают: Между группами – «Между вариантами», Внутри групп – «Остаточная», SS – сумма квадратов отклонений ($CKO = 2479,0$; $CKV=2445,5$; $CKE =33,5$), df – степени свободы, MS – средний квадрат отклонений или дисперсия – для вариантов $S_v^2 = 611,38$; остаточная дисперсия – $S_e^2 = 2,23$.

11. $F_\phi = 273,75$; $F_{05} = 3,05$. Так как $F_\phi > F_{05}$, $H_0 \neq 0$, нулевая гипотеза отвергается – в опыте в целом есть существенные различия, поэтому необходимо рассчитать наименьшую существенную разность – HCP .

В Пакете анализа программы Excel, а также в большинстве других статистических пакетов не предусмотрена оценка существенности средних по HCP – четвертый этап дисперсионного анализа, поэтому ошибку разности (S_d) и (HCP_{05}) можно рассчитать для нашего примера следующим образом:

$$S_d = \sqrt{\frac{2S_e^2}{n}} = \sqrt{\frac{2 \cdot 2,23}{4}} = 1,12 \quad HCP_{05} = t_{05} \cdot S_d = 2,13 \cdot 1,12 = 2,38 \text{ мг/г}$$

$t_{05} = 2,13$ при $df(cce) = 15$ степенях свободы для остатка

12. Составляем итоговую таблицу, в которую заносим средние значения по вариантам, рассчитываем отклонения между ними, по величине HCP определяем существенные различия между вариантами. Так как все отклонения по модулю больше HCP_{05} ($|d| > HCP_{05}$), значит, между опытными вариантами и контролем есть существенные различия на 05%-ном уровне значимости.

Биологический вывод: При выращивании растений на гормональных средах, дополненных одновременно ауксинами (ИУК, ИМК) и цитокининами (Кин, БАП) отмечается существенное снижение содержание СРФС по сравнению с таковыми у растений, выращенных на безгормональной среде. В то время как в действии цитокининов по ИУК нет существенных различий ($d = 1,25 < HCP_{05} = 2,38$).

Влияние сред разного гормонального состава на содержание суммы растворимых фенольных соединений (СРФС) в листьях клюквы крупноплодной, мг/г сырой массы

Варианты опыта	Сред- ние, \bar{x}_v	Отклонения $d = \bar{x}_{it} - \bar{x}_{st}$
1. Без гормонов – st	53,25	–
2. ИУК/Кин	33,25	–20,00
3. ИУК/БАП	34,50	–18,75
4. ИМК/Кин	21,00	–32,25
5. ИМК/БАП	25,50	–27,75

$HCP_{05} = 2,38$ мг/г

Работа 9. Дисперсионный анализ данных однофакторных экспериментов с рандомизированными блоками (повторениями)

Теоретическая часть:

В опытах, проведенных методом организованных (рандомизированных) блоков (повторений), общая изменчивость результативного признака с помощью дисперсионного анализа разлагается на три части: *варьирование повторений (блоков)*, *варьирование изучаемых вариантов* и *случайное*.

Так, в полевых опытах пестрота почвенного плодородия бывает внутри повторений и между повторениями. Изменчивость плодородия почвы, обусловленная влиянием систематического (закономерного) варьирования плодородия почвы рассчитывается, если варианты в полевом опыте сгруппированы по блокам или повторениям.

Изменчивость результативного признака, обусловленная пестротой почвенного плодородия внутри повторений, относится к случайным ошибкам.

Контрольные вопросы к защите:

1. Какие эксперименты относятся к опытам с зависимыми выборками?	3. Как размещаются варианты в опытах с рандомизированными блоками (повторениями)?
2. Схема (модель) дисперсионного анализа данных опыта, проведенного методом организованных повторений.	4. Группировка сортов (вариантов) по HSP_{05}

Общая постановка задачи:

- изучить дисперсионный анализ данных однофакторного опыта, проведенного методом организованных (рандомизированных) повторений;
- сделать статистический и биологический вывод.

Список индивидуальных заданий:

Таблица 35

Шифр заданий к работе 9

№	Повторность	№	Повторность	№	Повторность	№	Повторность
1	1 2 3 4	9	2 3 5 7	17	1 3 5 7	25	1 3 4 7
2	2 3 4 5	10	1 2 4 7	18	2 5 6 7	26	1 2 6 7
3	3 4 5 6	11	2 4 5 6	19	1 2 5 6	27	2 3 6 7
4	1 2 3 5	12	1 4 5 6	20	1 3 4 5	28	2 3 4 7
5	4 5 6 7	13	3 5 6 7	21	1 4 5 7	29	3 4 5 7
6	1 2 3 6	14	1 3 6 7	22	1 4 6 7	30	2 4 5 7
7	2 3 5 6	15	1 3 5 6	23	1 5 6 7	31	1 3 4 6
8	1 2 3 7	16	2 4 6 7	24	1 4 5 7	32	2 4 7

Пример. Влияние разных доз биотехнологического препарата Ризобакт СП на урожайность картофеля, т/га

Таблица 36

Варианты опыта	Повторения						
	I	II	III	IV	V	VI	VII
1. Без препарата (st)	14,0	15,2	14,9	13,9	14,2	14,3	14,9
2. Ризобакт -0,5 л/т	14,8	15,8	13,2	14,2	14,1	14,3	14,7
3. Ризобакт- 1,0 л/т	14,6	16,3	15,4	15,3	15,1	16,2	15,3
4. Ризобакт- 1,5 л/т	15,8	17,2	18,2	17,7	16,8	18,5	17,3
5. Ризобакт- 2,5 л/т	19,3	19,8	18,9	20,1	20,4	20,4	21,1

Задание № _____

Таблица 37

Исходная таблица

Варианты опыта	Повторения				Суммы V	Средние \bar{x}
	I	II	III	IV		
1.						
2.						
3.						
4.						
5.						
Суммы II						

$$\Sigma X = \Sigma V = \Sigma II =$$

Вспомогательная таблица для вычисления сумм квадратов отклонений

Варианты опыта	X^2				V^2
	I	II	III	IV	
1.					
2.					
3					
4					
5					
Π^2					—
	$\sum X^2 =$	$\sum V^2 =$	$\sum \Pi^2 =$		

Общее число наблюдений $N = v \cdot n =$

$$\text{Поправка } C = \frac{(\sum X)^2}{N} =$$

Суммы квадратов отклонений:

$$CKO = \sum X^2 - C =$$

$$CKV = \frac{\sum V^2}{n} - C =$$

$$CK\Pi = \frac{\sum \Pi^2}{v} - C =$$

$$CKE = CKO - CKV - CK\Pi =$$

Таблица 39

Таблица дисперсионного анализа

Источник вариации	Суммы квадратов	Доля вариации	Степени свободы	Средний квадрат	Критерий F	
					факт.	05
Общая				—	—	—
Повторений				—	—	—
Вариантов						
Остаточная					—	—

$$S_x = \sqrt{\frac{S_e^2}{n}} =$$

$$S_d = \sqrt{\frac{2S_e^2}{n}} =$$

$$HCP_{05} = t_{05} \cdot S_d =$$

$$HCP_{01} = t_{01} \cdot S_d =$$

Таблица сравнения средних

Варианты опыта	Средние, \bar{x}_v	Отклонения $d = \bar{x}_{on} - \bar{x}_{st}$	Группа
1 (st)			
2			
3			
4			
5			

 $HCP_{05} =$ $HCP_{01} =$

Вывод:

Оценка результатов:

Компетенции	Владение теорией	Владение практикой	Балл
Владеет методами оценки результатов полевого опыта на основе дисперсионного анализа	+	+	2
Владеет на удовлетворительно	+-	+-	1
Не владеет	-	-	0

Дата _____ Подпись преподавателя _____

Дисперсионный анализ данных однофакторных экспериментов с рандомизированными блоками (повторениями) в программе Excel

Пример. Влияние разных доз биотехнологического препарата Ризобакт СП на урожайность картофеля, т/га

Варианты опыта	Повторения			
	I	II	III	IV
1. Без препарата	15,2	13,9	14,2	14,9
2. Ризобакт – 0,5 л/т	15,8	14,2	14,1	14,7
3. Ризобакт – 1,0 л/т	16,3	15,3	15,1	15,3
4. Ризобакт – 1,5 л/т	17,2	17,7	16,8	17,3
5. Ризобакт – 2,0 л/т	19,8	20,1	20,4	21,1

1. В активный лист программы Excel введем исходные данные однофакторного полевого опыта, расположив таблицу в следующем виде (рис.9.1):

	A	B	C	D	E
1	Варианты	Повторения			
2		I	II	III	IV
3	1. Без препарата	15,2	13,9	14,2	14,9
4	2. Ризобакт – 0,5 л/т	15,8	14,2	14,1	14,7
5	3. Ризобакт – 1,0 л/т	16,3	15,3	15,1	15,3
6	4. Ризобакт – 1,5 л/т	17,2	17,7	16,8	17,3
7	5. Ризобакт – 2,0 л/т	19,8	20,1	20,4	21,1
8					
9					

Рис. 9.1. Исходные данные и Инструменты анализа данных

2. Дисперсионный анализ однофакторного полевого опыта с рандомизированными повторениями в программе Excel можно выполнить с помощью инструмента, который называется «Двухфакторный дисперсионный анализ без повторений». Из Пакета анализа выберем инструмент **Двухфакторный дисперсионный анализ без повторений** (рис. 9.1).

3. В появившемся окне укажем входной интервал A2:E7. Входной интервал должен включать только диапазон, состоящий из перечня вариантов, кодов повторений и цифровых данных (рис.9.2).

	A	B	C	D	E
1	Варианты	Повторения			
2		I	II	III	IV
3	1. Без препарата	15,2	13,9	14,2	14,9
4	2. Ризобакт – 0,5 л/т	15,8	14,2	14,1	14,7
5	3. Ризобакт – 1,0 л/т	16,3	15,3	15,1	15,3
6	4. Ризобакт – 1,5 л/т	17,2	17,7	16,8	17,3
7	5. Ризобакт – 2,0 л/т	19,8	20,1	20,4	21,1
8					
9					

Рис. 9.2. Диалоговое окно *Двухфакторный дисперсионный анализ без повторений*

4. Укажем метки в первом столбце (рис.9.2). Это необходимо для того, чтобы в выходных таблицах автоматически указывались наименования вариантов.

5. Альфа – выбор уровня значимости **0,05** или **0,01**

6. Выбираем выходной интервал для размещения результатов дисперсионного анализа: на данном листе или новом листе и нажимаем **ОК** (рис. 9.2.)

7. В итоге получаем таблицу дисперсионного анализа «Двухфакторный дисперсионный анализ без повторений», которая по сути является таблицей дисперсионного анализа данных нашего однофакторного полевого опыта, заложенного методом рандомизированных повторений в 4-х кратной повторности (рис.9. 3.)

Двухфакторный дисперсионный анализ без повторений						
ИТОГИ	Счет	Сумма	Среднее	Дисперсия		
1. Без препарата	4	58,2	14,55	0,36		
2. Ризобакт – 0,5 л/т	4	58,8	14,7	0,61		
3. Ризобакт – 1,0 л/т	4	62	15,5	0,29		
4. Ризобакт – 1,5 л/т	4	69	17,25	0,14		
5. Ризобакт – 2,0 л/т	4	81,4	20,35	0,31		
I	5	84,3	16,86	3,24		
II	5	81,2	16,24	6,89		
III	5	80,6	16,12	6,90		
IV	5	83,3	16,66	7,23		
Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критич
Строки	93,69	4	23,42	84,87	0,00	3,26
Столбцы	1,82	3	0,61	2,20	0,14	3,49
Погрешность	3,31	12	0,28			
Итого	98,82	19				

9.3. Итоговая таблица дисперсионного анализа

8. В верхней части итоговой части таблицы приведены суммы, средние значения и дисперсии по вариантам и повторениям (римские цифры), счет – это повторность каждого варианта ($n = 4$).

9. В нижней части представлена таблица дисперсионного анализа, в которой термины и формулы подразумевают: Строки – «Варианты», Столбцы – «Повторения», Погрешность – «Остаток», SS – сумма квадратов отклонений (CKO или C_y), df – степени свободы, MS – средний квадрат отклонений или дисперсия – S^2 .

10. $F_{\phi} = 84,87$ $F_{05} = 3,26$, так как $F_{\phi} > F_{05}$, и $p < 0,05$, $H_0 \neq 0$, нулевая гипотеза отвергается – в опыте в целом есть существенные различия, поэтому необходимо рассчитать HCP .

В Пакете анализа программы Excel, а также в большинстве других статистических пакетов не предусмотрена оценка существенности средних по HCP – четвертый этап дисперсионного анализа, поэтому ошибку разности (S_d) и (HCP_{05}) можно рассчитать для нашего примера следующим образом:

$$S_d = \sqrt{\frac{2S_e^2}{n}} = \sqrt{\frac{2 \cdot 0,28}{4}} = 0,37 \quad HCP_{05} = t_{05} \cdot S_d = 2,18 \cdot 0,37 = 0,81 \text{ т/га}$$

$t_{05} = 2,18$ при $df(cce) = 12$ степенях свободы для остатка

11. Составляем итоговую таблицу, в которую заносим средние значения по вариантам, рассчитываем отклонения (разности) между средними, и по величине HCP определяем существенные различия между вариантами.

Урожайность картофеля
в зависимости от разных доз биопрепарата, т/га

Варианты опыта	Сред- ние, \bar{x}_v	Отклонения $d = \bar{x}_{it} - \bar{x}_{st}$
1. Без препарата	14,55	–
2. Ризобакт – 0,5 л/т	14,70	0,15
3. Ризобакт – 1,0 л/т	15,50	0,95
4. Ризобакт – 1,5 л/т	17,25	2,70
5. Ризобакт – 2,0 л/т	20,35	5,80

$$HCP_{05} = 0,81 \text{ т/га}$$

Для определения существенных различий между средними значениями проведем сравнение отклонений с величиной наименьшей существенной разности (HCP_{05}). Так, в нашем опыте обработка клубней картофеля биотехнологическим препаратом Ризобакт в дозах 1,0 л/т и выше приводит к существенной прибавке урожайности картофеля по сравнению с контролем ($d > HCP_{05}$). Применение Ризобакта в дозе 0,5 л/т дает незначимую прибавку урожайности картофеля по сравнению с контролем, так как прибавка урожайности 0,15 т/га меньше HCP_{05} .

Работа 10. Дисперсионный анализ данных с неоднородными выборками
(анализ данных наблюдений, которые не подчиняются закону нормального распределения)

Контрольные вопросы:

1. Предпосылки дисперсионного анализа.	2. Способы преобразования исходных дат.
--	---

В соответствии с номером задания впишите в таблицу исходных данных результаты наблюдений по повторениям, проведите необходимые преобразования, уясните сущность дисперсионного анализа, составьте итоговую таблицу и сделайте вывод.

Таблица 41

Шифр задания к работе 10

№	Повторность	№	Повторность	№	Повторность	№	Повторность
1	1 2 3 4	9	2 3 5 7	17	1 3 5 7	25	1 3 4 7
2	2 3 4 5	10	1 2 4 7	18	2 5 6 7	26	1 2 6 7
3	3 4 5 6	11	2 4 5 6	19	1 2 5 6	27	2 3 6 7
4	1 2 3 5	12	1 4 5 6	20	1 3 4 5	28	2 3 4 7
5	4 5 6 7	13	3 5 6 7	21	1 4 5 7	29	3 4 5 7
6	1 2 3 6	14	1 3 6 7	22	1 4 6 7	30	2 4 5 7
7	2 3 5 6	15	1 3 5 6	23	1 5 6 7	31	1 3 4 6
8	1 2 3 7	16	2 4 6 7	24	1 4 5 7	32	2 4 7

Пример 1. Количество растений на однорядковой делянке 2,5 м с отклонениями от нормы (карлики, скороспелые)

Таблица 42

Варианты опыта	Повторения						
	1	2	3	4	5	6	7
1. Диметилсульфат	2	2	1	0	4	0	2
2. Этиленэмин	2	3	4	4	3	4	7
3. Нитрозэтилмочевина	6	7	8	12	8	13	9
4. Этилметансульфонат	9	13	28	17	25	18	27
5. Смесь мутагенов	37	25	37	29	37	28	19

Пример 2. Влияние гиббереллина на пораженность картофеля морщинистой мозаикой, %

Таблица 43

Варианты опыта	Повторения						
	1	2	3	4	5	6	7
1. Без обработки	70	72	69	68	65	72	75
2. 0,0001 %	58	62	60	64	65	55	64
3. 0,0005 %	40	43	46	45	47	50	49
4. 0,001 %	15	11	12	18	16	14	16

Решение:

задание № _____

Таблица 44

Варианты опыта	Повторения				Суммы V	Средние \bar{x}_v
	1	2	3	4		
1.						
2.						
3.						
4.						
5.						
Суммы P					$\sum X =$	

Таблица 45

Преобразование (трансформация исходных дат) $X_I =$

Варианты опыта	Повторения, X_I				Суммы V	Средние \bar{x}_v
	1	2	3	4		
1.						
2.						
3.						
4.						
5.						
Суммы P					$\sum X_I =$	

$$\sum X_I = \sum V = \sum P$$

Таблица 46

Вспомогательная таблица для вычисления сумм квадратов отклонений

Варианты опыта	X^2				V^2
	I	II	III	IV	
1.					
2.					
3.					
4.					
5.					
Π^2					—

$$\sum X^2 =$$

$$\sum V^2 =$$

$$\sum \Pi^2 =$$

Общее число наблюдений $N = v \cdot n =$

$$\text{Поправка } C = \frac{(\sum X)^2}{N} =$$

Суммы квадратов отклонений:

$$CKO = \sum X^2 - C =$$

$$CKV = \frac{\sum V^2}{n} - C =$$

$$CKП = \frac{\sum P^2}{v} - C =$$

$$CKE = CKO - CKV - CKП =$$

Таблица 47

Таблица дисперсионного анализа

Виды варьирования	Суммы квадратов	Степени свободы	Средний квадрат	Критерий F	
				факт.	05
Общее			—	—	—
Повторений			—	—	—
Вариантов					
Остаток				—	—

$$S_d = \sqrt{\frac{2S_e^2}{n}} =$$

$$HCP_{05} = t_{05} \cdot S_d =$$

$$HCP_{01} = t_{01} \cdot S_d =$$

Таблица 48

Итоговая таблица

Варианты опыта	Средние преобразов., \bar{x}_{v1}	$d = \bar{x}_{on} - \bar{x}_{st}$	Средние исходные, \bar{x}_v	$d = \bar{x}_{on} - \bar{x}_{st}$
1.				
2..				
3.				
4.				
5.				

$$HCP_{05} =$$

$$HCP_{01} =$$

После оценки существенности разности преобразованных значений, результаты оценки перенести на исходные средние. Существенные разности обозначьте звездочками (*)

Вывод:

Оценка результатов

Компетенции	Владение теорией	Владение практикой	Балл
Владеет методами обработки данных с неоднородными выборками	+	+	2
Владеет на удовлетворительно	+-	+-	1
Не владеет	-	-	0

Дата _____ Подпись преподавателя _____

Работа 11. Дисперсионный анализ данных двухфакторного опыта

Теоретическая часть:

Многофакторный дисперсионный комплекс – это совокупность исходных наблюдений (дат), позволяющих статистически оценить действие и взаимодействие изучаемых факторов на изменчивость результативного признака. Эффект взаимодействия составляет ту часть общую варьирования, которая вызвана различным действием одного фактора при разных градациях другого.

Задачей многофакторного опыта является изучение влияния факторов в отдельности и их взаимодействия на результативный признак. В многофакторном опыте вариантами являются градации нескольких факторов, взятые в отдельности и в сочетаниях. Поэтому изменчивость по вариантам включает в себя изменчивость, обусловленную каждым фактором в отдельности и их взаимодействием. Так в двухфакторном опыте сумма квадратов для вариантов расчленяется на три компонента: на сумму квадратов для фактора A , сумму квадратов для фактора B и сумму квадратов для их взаимодействия AB .

Статистическую обработку урожайных данных двухфакторного опыта, проведенного методом полной рандомизации, проводят в следующей последовательности:

- 1) исходные даты заносят в таблицу урожаев, определяют суммы и средние;
- 2) вычисляют суммы квадратов для общего варьирования, вариантов и остатка (ошибки), т.е. обрабатывают данные так же, как и результаты однофакторного опыта;
- 3) сумму квадратов вариантов расчленяют на компоненты: главные эффекты изучаемых факторов и их взаимодействия;
- 4) составляют итоговую таблицу дисперсионного анализа и проверяют нулевую гипотезу о существенности действия и взаимодействия факторов по критерию Фишера;
- 5) рассчитывают HCP для оценки частных различий и HCP для оценки главных эффектов и их взаимодействия.

Контрольные вопросы к защите:

1. Особенности многофакторных опытов. Их краткое обозначение.	3. Различия в моделях ДА двухфакторного опыта, проведенного методом организованных повторений и полной рандомизации
2. Матрица планирования 2-х факторного опыта 5x2	4. Как осуществляется оценка средних значений по вариантам в многофакторных опытах?

Общая постановка задачи:

- изучить дисперсионный анализ данных многофакторного полевого опыта, проведенного методом полной рандомизации;
- освоить практические навыки проведения дисперсионного анализа многофакторных полевых опытов;
- сделать статистический и биологический вывод

Список индивидуальных заданий:

Вариант	<i>A</i>	<i>B</i>	<i>n</i>
1	0 1 2	0 1 2 3	4
2	0 1 2	0 1 2	4
3	0 1 2	0 1 3	4
4	0 1 2	0 2 3	4
5	0 1 2	1 2 3	4
6	0 1 2	0 1 2 3	3
7	0 1 2	0 1 2	3
8	0 1 2	0 1 3	3
9	0 1 2	0 2 3	3
10	0 1 2	1 2 3	3

Пример 1. Дисперсионный анализ двухфакторного полевого опыта 3x4, заложенного методом полной рандомизации.

В опыте изучается влияние стимуляторов роста и доз азота на урожайность озимой пшеницы, т/га.

Фактор А – стимуляторы роста (**0** – без стимулятора, контроль, **1** – Эпин, **2** – Мивал-агро)

Фактор В – 4 дозы азота, кг/га (**0** – без азота, **1** – 60, **2** – 90, **3** – 120)

Таблица 49

Фактор А	Фактор В	Повторности				Суммы, V	Средние \bar{x}_v
		I	II	III	IV		
0	0	3,0	3,2	3,3	3,1		
	1	5,2	5,3	5,4	5,6		
	2	5,8	5,9	5,9	5,7		
	3	6,0	6,0	6,2	6,4		
1	0	3,4	3,6	3,5	3,4		
	1	6,0	6,3	6,2	6,5		
	2	6,7	6,8	6,5	6,9		
	3	6,3	5,9	6,5	6,8		
2	0	3,8	3,9	3,7	3,6		
	1	7,0	6,9	6,8	6,7		
	2	6,4	6,6	6,2	6,4		
	3	6,7	6,5	6,9	6,8		
Суммы П						ΣX	

Дисперсионный анализ по модели однофакторного опыта – ДА-1:

Общее число наблюдений $N = a \cdot b \cdot n =$

где a и b – число градаций фактора A и B , $v = a \cdot b$

$$\text{Поправка } C = \frac{(\sum X)^2}{N} =$$

Суммы квадратов отклонений (СК):

$$\text{общая } SKO = \sum X^2 - C =$$

$$\text{вариантов } SKV = \frac{\sum V^2}{n} - C =$$

$$\text{остаток } SKE = SKO - SKV =$$

Степени свободы (df):

$$N - 1 =$$

$$v - 1 = (a \cdot b - 1) =$$

$$N - v =$$

Таблица 50

Таблица дисперсионного анализа (ДА-1)

Источник вариации	Суммы квадратов	Степени свободы	Средний квадрат	Критерий F	
				факт.	05
Общая			–	–	–
Вариантов					
Остаточная				–	–

По критерию Фишера определяем существенность различий в опыте и при $F_{ф} \geq F_{05}$ расчлняем сумму квадратов по вариантам $СКV$ на составляющие ее виды вариации.

ДА-2. Определение сумм квадратов для факторов A , B и взаимодействия AB

Исходную таблицу ДА-2 готовят на основе сумм по вариантам (V)

A (стимуляторы роста)	B (дозы азота)				Суммы A
	0	1	2	3	
0					
1					
2					
Суммы B					$\Sigma X =$

Суммы квадратов:

вариантов $СК_{(A+B+AB)} = СКV =$ (получена в ДА-1)

фактора A $СК_A = \frac{\sum A}{b \cdot n} - C =$ $df = a - 1 =$

фактора B $СК_B = \frac{\sum B}{a \cdot n} - C =$ $df = b - 1 =$

взаимодействия $СК_{AB} = СКV - СК_A - СК_B =$ $df = (a - 1)(b - 1) =$

Результаты 2-х дисперсионных анализов сводят в сводную таблицу и определяют существенность действия и взаимодействия факторов по F-критерию

Таблица 51

Сводная таблица дисперсионного анализа

Источник вариации	Суммы квадратов	Доля вариации	Степени свободы	Средний квадрат	Критерий F	
					факт.	05
Общая		100%		–	–	–
Фактор A						
Фактор B						
Взаимодействие AB						
Остаток					–	–

Оценка существенности частных различий:

$$S_{\bar{x}} = \sqrt{\frac{S_e^2}{n}}$$

$$S_d = \sqrt{\frac{2S_e^2}{n}} =;$$

$$t_{05}(ccE_1) =$$

$$HCP_{05} = t_{05}S =$$

Оценка существенности главных эффектов и взаимодействия:
стимуляторы роста (A)

$$S_d^A = \sqrt{\frac{2S_e^2}{nb}} =$$

$$HCP_{05}^A = t_{05}S_d^A =$$

дозы азота (B) и взаимодействие (AB):

$$S_d^B = \sqrt{\frac{2S_e^2}{na}} =$$

$$HCP_{05}^B = t_{05}S_d^B =$$

Таблица 52

Влияние стимуляторов роста и доз азота на урожайность озимой пшеницы, т/га

A (стимуляторы роста)	B (дозы азота, кг/га)				Средние по A $HCP_{05}^A =$
	0	60	90	120	
Контроль					
Эпин					
Мивал-агро					
Средние по B $HCP_{05}^B =$					

$$HCP_{05} =$$

Представляем итоговые данные графически:

т/га



Дозы азота, кг/га

Рис. 4. Урожайность озимой пшеницы в зависимости от стимуляторов роста и доз азота

Вывод:

Оценка результатов:

Компетенции	Владение теорией	Владение практикой	Балл
Владеет методами планирования многофакторных опытов на основе дисперсионного анализа	+	+	2
Владеет на удовлетворительно	+-	+-	1
Не владеет	-	-	0

Дата _____ Подпись преподавателя _____

Дисперсионный анализ данных двухфакторного опыта с полной рандомизацией вариантов в программе Excel

Пример1. В двухфакторном опыте 2x3, проведенном методом полной рандомизации (независимые выборки) изучается два фактора: фактор А – Стимуляторы в 2-х градациях: (a_0 – без стимулятора, a_2 – Эпин), фактор В – дозы азота: ($b_1 = 0$ $b_2 = 60$ $b_3 = 90$) на урожайность озимой пшеницы. Опыт проведен в 4-х кратной повторности ($n=4$), ц/га

Стимуляторы А	Дозы азота В	Повторность			
		<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>
<i>Без стимулятора</i>	<i>0</i>	24,1	25,8	23,0	27,0
	<i>60</i>	28,4	29,7	30,1	27,4
	<i>90</i>	28,7	30,4	32,0	27,0
<i>Эпин</i>	<i>0</i>	30,7	34,4	34,0	31,0
	<i>60</i>	46,7	45,4	47,1	46,3
	<i>90</i>	59,4	50,7	64,5	60,1

1. В активный лист программы Excel введем исходные данные вышеприведенного примера, расположив таблицу в следующем виде (рис.11.1):

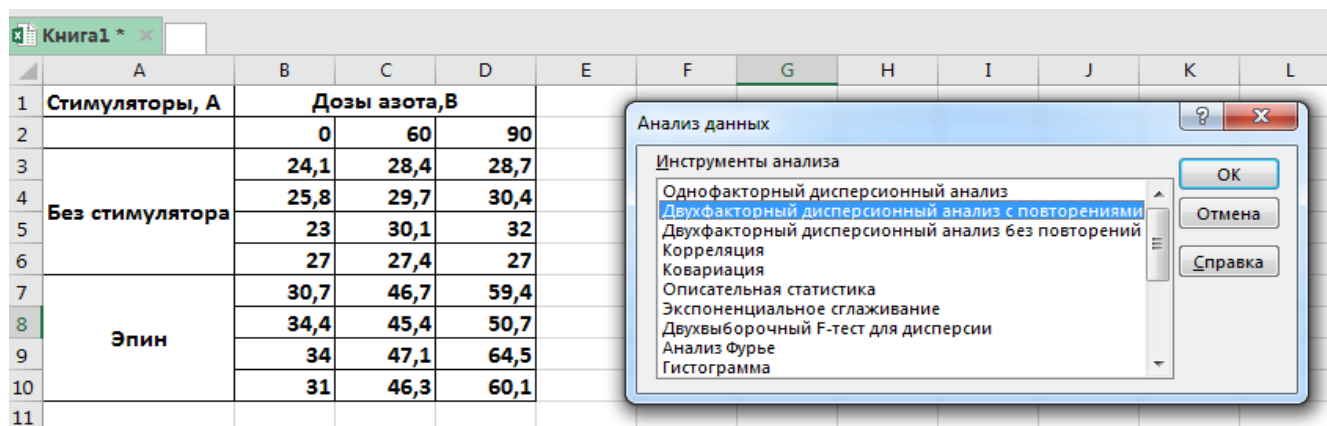


Рис. 11.1. Исходные данные и Инструменты анализа данных

2. Из Пакета анализа выберем инструмент **Двухфакторный дисперсионный анализ с повторениями**, (рис. 11.1).

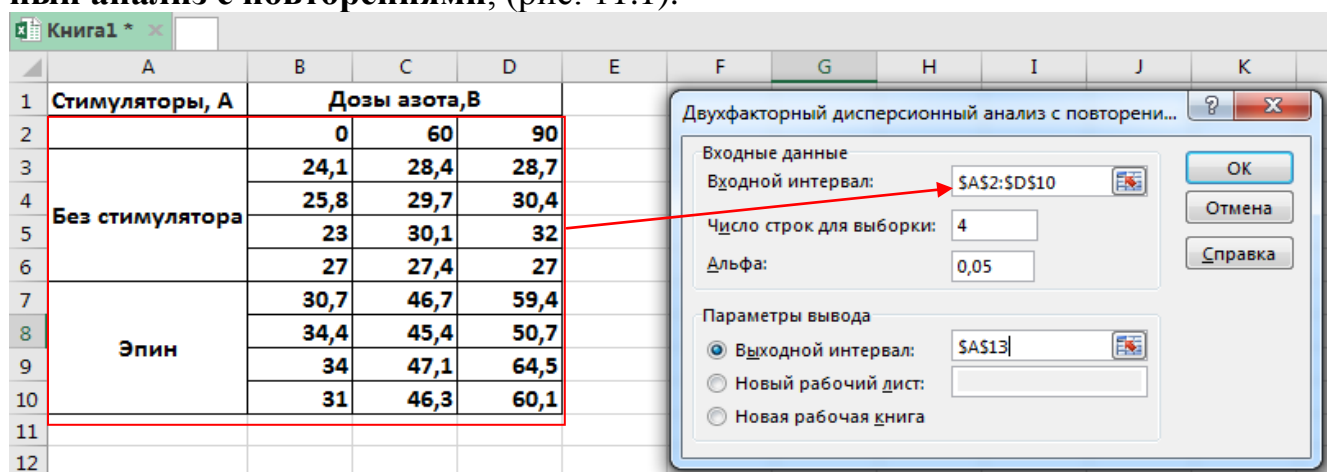


Рис.11.2. Диалоговое окно **Двухфакторный дисперсионный анализ с повторениями**

3. В появившемся окне укажем входной интервал **A2:D10**. Входной интервал должен включать диапазон, состоящий из перечня вариантов по фактору А, фактору В и цифровых данных по этим вариантам (рис.11. 2).

4. В окне Число строк для выборки – укажем **4** (это повторность опыта) (рис.11.2).

5. Альфа – выбор уровня значимости **0,05** или **0,01**

6. Выберем выходной интервал для размещения результатов дисперсионного анализа: выберем на этом или на новом листе и нажмем на клавишу **ОК** (рис.11.2.)

7. Получаем итоговую таблицу дисперсионного анализа «Двухфакторный дисперсионный анализ с повторениями» (рис. 11.3.)

Двухфакторный дисперсионный анализ с повторениями						
ИТОГИ	0	60	90	Итого		
Без стимулятора						
Счет	4	4	4	12		
Сумма	99,90	115,60	118,10	333,60		
Среднее	24,98	28,90	29,53	27,80		
Дисперсия	3,15	1,53	4,65	6,97		
Эпин						
Счет	4	4	4	12		
Сумма	130,10	185,50	234,70	550,30		
Среднее	32,53	46,38	58,68	45,86		
Дисперсия	3,78	0,53	33,36	134,75		
Итого						
Счет	8	8	8			
Сумма	230,00	301,10	352,80			
Среднее	28,75	37,64	44,10			
Дисперсия	19,26	88,13	259,07			
Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критич
Выборка	1956,62	1	1956,62	249,79	0,00	4,41
Столбцы	950,33	2	475,17	60,66	0,00	3,55
Взаимодействие	467,58	2	233,79	29,85	0,00	3,55
Внутри	141,00	18	7,83			
Итого	3515,53	23				

Рис.11. 3. Итоговая таблица дисперсионного анализа данных двухфакторного опыта

9. В верхней итоговой таблице Excel представлены суммы и средние значения по факторам А и В, которые удобнее представить в виде обобщенной таблицы средних по изучаемым вариантам. Счет – это повторность каждого варианта ($n = 4$).

10. В нижней итоговой части представлена стандартная таблица дисперсионного анализа, в которой термины и формулы подразумевают следующее: Выборка – «Фактор А», Столбцы – «Фактор В», Взаимодействие – «Взаимодействие АВ», Внутри – «Остаток». SS – сумма квадратов отклонений (CKO или C_y), df – степени свободы, MS – средний квадрат отклонений или дисперсия – S^2 .

11. С помощью двухфакторного дисперсионного анализа по критерию Фишера оценивается отдельно существенность изучаемых факторов и их взаимодействия. В нашем примере для фактора А $F_\phi = 249,78$; $F_{05} = 4,41$, для фактора В $F_\phi = 60,66$; $F_{05} = 3,55$, для взаимодействия АВ $F_\phi = 29,85$; $F_{05} = 3,55$. Так как $F_\phi > F_{05}$, $H_0 \neq 0$, нулевая гипотеза отвергается – действие и взаимодействие удобрений и

стимуляторов роста значимо на 5% ном уровне значимости. Для оценки существенности разности средних необходимо рассчитать величину HCP_{05} .

$$S_d = \sqrt{\frac{2S_e^2}{n}} = \sqrt{\frac{2 \cdot 7,83}{4}} = 1,98 \quad HCP_{05} = t_{05} \cdot S_d = 2,10 \cdot 1,98 = 4,15 \text{ ц/га.}$$

$t_{05} = 2,10$ при $df(cce) = 18$ степенях свободы для остатка

$$S_d^A = \sqrt{\frac{2S_e^2}{n \cdot b}} = \sqrt{\frac{2 \cdot 7,83}{4 \cdot 3}} = 1,14 \quad HCP_{05}^A = t_{05} \cdot S_d^A = 2,10 \cdot 1,14 = 2,39 \text{ ц/га.}$$

$$S_d^B = \sqrt{\frac{2S_e^2}{n \cdot a}} = \sqrt{\frac{2 \cdot 7,83}{4 \cdot 2}} = 1,40 \quad HCP_{05}^B = t_{05} \cdot S_d^B = 2,10 \cdot 1,39 = 2,93 \text{ ц/га.}$$

Итоговая таблица

Влияние стимуляторов роста и доз азота
на урожайность озимой пшеницы, ц/га

Фактор А – стимуляторы	Фактор В – дозы азота			В среднем по фактору А $HCP_{05}^A = 2,39$
	0	60	90	
Без стимулятора	24,98	28,90	29,53	27,80
Эпин	32,53	46,38	58,68	45,86
В среднем по фактору В $HCP_{05}^B = 2,93$	28,76	37,64	44,10	

$$HCP_{05} = 4,15$$

С помощью $HCP_{05} = 4,15$ оцениваются различия между частными средними (с эпином и без стимулятора при разных дозах азота: 28,90 – 24,98; 58,68 – 29,53; 46,38 – 32,53 и т.д.).

$HCP_{05}^A = 2,39$ оценивает только главный эффект фактора А (45,86 – 27,80), а $HCP_{05}^B = 2,93$ различия главного эффекта фактора В (44,10 – 28,76; 37,64 – 28,76; 44,10 – 37,64).

Работа 12. Планирование биотехнологического опыта (эксперимента)

Планирование классического опыта (эксперимента)

Методические указания к выполнению работы:

1. Выбор темы. Формулировка темы опыта должна быть из ключевых слов, отражающих суть изучаемого вопроса. Из названия темы должно быть понятно, какая проблема поставлена в научном исследовании.

Для решения этого вопроса можно воспользоваться тематическим планом исследований кафедры, ранее выполненными исследованиями, но на новом методическом уровне. Можно спланировать эксперимент по теме будущей дипломной работы.

2. Выдвижение рабочей гипотезы, т.е. научное предположение о действии изучаемых в опыте вариантов. Рабочая гипотеза выдвигается на основании изучения литературы по выбранной теме. Рабочая гипотеза должна быть краткой, обладать логичностью и быть приложимой к достаточно широкому кругу явлений. В рабочей гипотезе обычно отражается, во-первых, что произойдет, если будет произведено данное действие, а во-вторых, почему это произойдет именно так, а не иначе.

3. Определение цели и задач исследования. В любом опыте может быть только одна цель, а задач – несколько. Цель исследования – это общая его направленность на конечный результат. При формулировании цели исследования необходимо четко в одно предложение записать, что вы желаете добиться, изучить посредством данного опыта. Задачи исследования – это то, что требует решения в процессе исследования; вопросы, на которые должен быть получен ответ. Решением задач в опыте исследователь достигает поставленной цели. Задачи исследования – указать, каким путем будет достигаться поставленная цель, что необходимо проделать для этого. Обычно перечень задач содержит 2 – 4 пункта.

4. Разработка схемы опыта. Схема опыта с качественной градацией изучаемых вариантов должна быть спланирована таким образом, чтобы был соблюден принцип единственного различия и правильно выбран стандартный или контрольный вариант, а при планировании схемы опыта для вариантов с количественной градацией по результатам опыта можно было бы получить кривую отклика воздействия изучаемых вариантов на изучаемые признаки.

5. Разработка программы наблюдений и учетов. Наблюдения и учеты должны быть целенаправленными, то есть они должны соответствовать теме и схеме опыта. Необходимо указать периодичность проведения учетов, а также краткую методику их проведения.

Пример составления схемы опыта (с качественной градацией вариантов):

"Оценка действия новых биопрепаратов в посевах озимой пшеницы в условиях Московской области".

Рабочая гипотеза: Применение новых биопрепаратов обеспечивает более высокую (на 20-30%) урожайность и подавление грибных заболеваний озимой пшеницы.

Цель: Выявление наиболее эффективных биопрепаратов при выращивании озимой пшеницы в условиях Московской области.

Задачи: 1. Изучить действие новых биопрепаратов на пораженность озимой пшеницы болезнями (корневая гниль, мучнистая роса и т.д.)

2. Оценить влияние биопрепаратов на урожайность и качество урожая озимой пшеницы.

3. Дать сравнительную оценку технической (биологической), хозяйственной и экономической эффективности изучаемых биопрепаратов.

Схема опыта

1. Без биопрепаратов (контроль)
2. Азотобактерин (стандарт)
3. Флавобактерин
4. Мизорин
5. Ризоаргин
6. Биосолби Сан

Алгоритм планирования опыта:

Тема опыта _____

Рабочая гипотеза _____

Цель и задачи опыта _____

Схема опыта (перечень вариантов)

- 1.
- 2.
- 3.
- 4.
- 5.
- .

*Методы статистической обработки результатов опыта
(модель дисперсионного анализа данных опыта и т.д.)*

Программа наблюдений и учетов в планируемом опыте:

Краткая методика проведения опыта:

Планирование регрессионного многофакторного эксперимента

Изучение любой системы можно представить, как нахождение зависимостей вида:

$$Y=f(X_1, X_2, X_3 \dots X_k)$$

где $X_1, X_2 \dots, X_k$ – независимые переменные величины (факторы), влияющие на результативный признак (планируемый процесс).

При выборе факторов, влияющих на изучаемый биотехнологический процесс, необходимо учитывать следующие требования:

- факторы должны быть управляемыми, т.е. установлены на выбранных уровнях;
- факторы должны быть однозначными и непосредственно воздействовать на объект исследования;
- точность измерения факторов должна отвечать поставленной задаче;
- факторы должны быть совместимы, это означает, что все их комбинации должны быть осуществимы и безопасны;
- факторы должны быть независимы, т.е. каждый из факторов может быть установлен на любом уровне вне зависимости от уровней других факторов.

1. Сформулировать тему эксперимента (опыта)
2. Составить схему двухфакторного 2x2 или трехфакторного 2x2x2 опыта.
3. Составить матрицу полного 2-х или 3-х факторного эксперимента

Матрица планирования 2-х факторного опыта 2²

№	Факторы			Функция отклика	Обозначения
	A (X ₁)	B (X ₂)	C (X ₁)		
1					
2					
3					
4					
5					
6					
7					
8					
Σ	0	0	0		

Результаты ПФЭ 2^k
(для двухфакторного опыта 2², трехфакторного 2³)

№	Варианты опыта,			Y			Средняя \bar{y}	Дисперсия S_i^2
	A (X ₁)	B (X ₂)	C (X ₃)	Y ₁	Y ₂	Y ₃		
1								
2								
3								
4								
5								
6								
7								
8								

Определение воспроизводимости опыта

1. Расчет среднего значения $\bar{y}_i =$

2. Дисперсия для каждой серии $S_i^2 = \frac{\sum (Y_i - \bar{y}_i)^2}{n-1}$

Сумма дисперсий $\sum S_i^2 =$

Расчетное значение критерия Кохрена $G_p = \frac{\max S_i^2}{\sum S_i^2} =$

Табличное значение критерия Кохрена $G_{05} =$

при $df_1 = n - 1 =$ (числитель) и $df_2 = v =$ (знаменатель)

Если $G_p < G_{05}$, опыт воспроизводимый, дисперсии однородные

Оценка дисперсии воспроизводимости

$$S_y^2 = \frac{\sum S_i^2}{v} =$$

Оценка дисперсии среднего значения

$$S_{\bar{y}}^2 = \frac{S_y^2}{n} =$$

Проверка воспроизводимости опытов по критерию Кохрена G_{ϕ}

Опыт воспроизводим, если $G_{табл} \geq G_{\phi}$ дисперсии однородны и к таким опытам можно применять математические методы. Если опыт воспроизводим, можно провести математический анализ.

Построение уравнения регрессии

1. Математическое описание результатов эксперимента:

Уравнение регрессии: $Y = b_0 + b_1X_1 + b_2X_2 + b_{1,2}X_1X_2$

где:

Y – функция отклика, $b_0, b_1, b_2, b_{1,2}$ – коэффициенты регрессии

X_1, X_2 – кодированные переменные:

$$X_1 = \frac{A - A_{\min}}{\Delta a} \quad X_2 = \frac{B - B_{\min}}{\Delta b}$$

$$X_1 = \frac{A - A_{\min}}{\Delta a} = X_2 = \frac{B - B_{\min}}{\Delta b} =$$

2. Расчет коэффициентов регрессии:

$$b_0 = \frac{\sum \bar{y}_i}{v}; \quad b_1 = \frac{\sum X_1 \cdot \bar{y}_i}{v}; \quad b_2 = \frac{\sum X_2 \cdot \bar{y}_i}{v}; \quad b_{1,2} = \frac{\sum X_1 \cdot X_2 \cdot \bar{y}_i}{v}$$

Коэффициенты регрессии:

$$b_0 = \frac{\sum \bar{y}_i}{v} =$$

$$b_1 = \frac{\sum X_1 \cdot \bar{y}_i}{v} =$$

$$b_2 = \frac{\sum X_2 \cdot \bar{y}_i}{v} =$$

$$b_{1,2} = \frac{\sum X_1 \cdot X_2 \cdot \bar{y}_i}{v} =$$

$$Y = b_0 + b_1X_1 + b_2X_2 + b_{1,2}X_1X_2 =$$

3. Определение значимости коэффициентов регрессии

Ошибка коэффициента регрессии S_b

$$S_b = \sqrt{\frac{S_y^2}{v}} =$$

Критерий существенности $t_b = \frac{|\pm b_{yx}|}{S_b}$

Для $b_0 t_b = \frac{b_0}{S_b} = **$, для $b_1 t_b = \frac{b_1}{S_b} = **$

для $b_2 t_b = \frac{b_2}{S_b} = *$, для $b_{1,2} t_b = \frac{b_{1,2}}{S_b} =$

Критерий существенности коэффициента регрессии $t_{\text{теор}}, t_{05}, t_{01} =$ при $v \cdot n - 2 =$

3. Расчетные значения (Y_p) уравнения: $Y = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2$

$$Y_{p1} = b_0 + b_1(-1) + b_2(-1) + b_{12}(-1)(-1) =$$

$$Y_{p2} = b_0 + b_1(+1) + b_2(-1) + b_{12}(+1)(-1) =$$

$$Y_{p3} = b_0 + b_1(-1) + b_2(+1) + b_{12}(-1)(+1) =$$

$$Y_{p4} = b_0 + b_1(+1) + b_2(+1) + b_{12}(+1)(+1) =$$

4. Оценка дисперсии адекватности:

$$\text{Дисперсия адекватности: } S_{ae}^2 = \frac{\sum (Y_\phi - Y_p)^2}{(v-b)}$$

где: Y_ϕ – фактические значения, Y_p – расчетные значения показателя
 v – число вариантов, b – число оставшихся коэффициентов регрессии

5. Оценка адекватности уравнения регрессии по критерию Фишера

$$F_\phi = \frac{\max S_{ad}^2}{S_y^2} =$$

$F_{05} =$ при числе степеней свободы для числителя $df_1 = v-b =$

и степеней свободы для знаменателя $df_2 = v(n-1) =$

При $F_{05} > F_\phi$ уравнение $Y = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2$

адекватно описывает зависимость

Оценка результатов:

Компетенции	Владение теорией	Владение практикой	Балл
Освоил общие принципы планирования экспериментов	+	+	2
Освоил на удовлетворительно	+-	+-	1
Не владеет	-	-	0

Дата _____ Подпись преподавателя _____

ПРИЛОЖЕНИЯ. СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ.

Приложение 1

Таблица случайных чисел

i j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	4	7	5	8	6	3	5	1	9	1	6	1	7	4	2	0	2	3	5	9	3
2	1	9	3	5	3	4	9	5	2	7	5	8	4	3	4	8	1	2	7	0	1
3	2	8	7	1	4	1	0	2	4	1	5	2	9	4	6	2	1	5	9	8	7
4	8	4	8	5	1	3	9	6	7	2	1	9	2	4	6	7	3	0	5	1	3
5	3	8	4	7	5	0	5	1	7	3	4	8	2	6	0	4	3	6	1	2	1
6	3	5	3	1	9	3	7	4	9	0	6	2	1	4	6	2	5	4	5	8	0
7	3	4	5	9	5	2	7	9	8	9	5	8	5	1	7	3	5	1	7	2	9
8	4	1	5	3	9	1	3	7	2	5	8	7	1	3	6	4	9	2	8	7	9
9	7	2	9	5	6	7	8	5	4	5	3	4	3	4	1	9	8	6	2	1	0
10	5	9	2	8	9	8	6	4	1	5	3	7	4	3	2	5	0	6	1	2	3
11	1	3	4	3	9	6	5	2	8	7	4	9	3	7	3	2	8	0	4	5	2
12	4	6	2	1	3	8	6	2	0	5	3	1	5	4	9	3	7	4	9	7	5
13	7	8	6	3	2	0	5	1	9	5	6	4	3	5	9	1	8	7	4	2	4
14	3	8	2	7	5	0	4	3	4	5	7	9	6	9	5	1	7	6	8	5	1
15	6	2	8	9	4	5	3	7	2	0	4	1	5	6	3	5	1	4	6	7	6

Приложение 2

Теоретические значения критерия Стьюдента

Степени свободы, сс	Уровень значимости		Степени свободы, сс	Уровень значимости	
	05	01		05	01
<i>1</i>	12,71	63,66	<i>12</i>	2,18	3,06
<i>2</i>	4,30	9,93	<i>14</i>	2,15	2,98
<i>3</i>	3,18	5,84	<i>16</i>	2,12	2,92
<i>4</i>	2,78	4,60	<i>18</i>	2,10	2,88
<i>5</i>	2,57	4,03	<i>20</i>	2,09	2,85
<i>6</i>	2,45	3,71	<i>25</i>	2,06	2,79
<i>7</i>	2,37	3,50	<i>30</i>	2,04	2,75
<i>8</i>	2,31	3,36	<i>50</i>	2,01	2,68
<i>9</i>	2,26	3,25	<i>100</i>	1,98	2,63
<i>10</i>	2,23	3,17	∞	1,96	2,58

Теоретические значения критерия Фишера (F) (df_1 – число степеней свободы для дисперсии числителя (варианта), df_2 – знаменателя (остатка))

F_{05} – вероятность ошибки $\alpha = 0,05$ (05%)													
df_2	df_1												
	1	2	3	4	5	6	7	8	9	10	15	20	50
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,62	4,56	4,44
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,94	3,87	3,75
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,51	3,44	3,32
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,22	3,15	3,03
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,01	2,94	2,80
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,85	2,77	2,64
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,72	2,65	2,50
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,60	2,55	2,40
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,40	2,33	2,18
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,20	2,12	1,96
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,01	1,93	1,76
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,92	1,78	1,60
F_{01} – вероятность ошибки $\alpha = 0,01$ (01%)													
df_2	df_1												
	1	2	3	4	5	6	7	8	9	10	15	20	50
5	16,3	13,3	12,1	11,4	10,9	10,7	10,4	10,3	10,2	10,0	9,72	9,55	9,24
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,56	7,39	7,09
7	12,2	9,5	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,31	6,15	5,85
8	11,3	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,52	5,36	5,06
9	10,6	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	4,96	4,80	4,51
10	10,0	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,56	4,41	4,12
11	9,65	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,25	4,10	3,80
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,02	3,70	3,56
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,57	3,36	3,07
20	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37	3,09	2,94	2,63
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,06	2,98	2,70	2,55	2,24
50	7,17	5,06	4,20	3,72	3,41	3,18	3,02	2,88	2,78	2,70	2,42	2,26	1,94

Значение функции $\Psi \left[\frac{R}{n+1} \right]$

$\frac{R}{n+1}$	Ψ	$\frac{R}{n+1}$	Ψ	$\frac{R}{n+1}$	Ψ
0.00	∞	0.25	-0.67	0.50	0.00
0.02	-2.05	0.29	-0.56	0.62	0.31
0.05	-1.64	0.30	-0.53	0.70	0.52
0.10	-1.28	0.33	-0.44	0.80	0.84
0.15	-1.04	0.40	-0.25	0.90	1.28
0.20	-0.84	0.43	-0.17	0.99	2.33

Теоретические значения X- критерия Ван-дер-Вандена

$n=n_1+n_2$	$n_1 - n_2 = 0$ или 1		$n_1 - n_2 = 2$ или 3	
	0.05	0.01	0.05	0.01
8	2.40	—	2.30	—
10	2.60	3.20	2.49	3.10
12	2.86	3.60	2.79	3.58
14	3.11	3.94	3.06	3.88
16	3.39	4.26	3.36	4.25
18	3.63	4.60	3.60	4.58
20	3.86	4.94	3.84	4.92
30	4.88	6.35	4.87	6.34
50	6.50	8.51	6.51	8.50

Теоретические значения T- критерия Уайта на 5%-ом уровне значимости (данные или меньшие значения отвергают H_0)

n_2	n_1						
	4	5	6	7	8	9	10
5	11	17					
6	12	18	26				
7	13	20	27	36			
8	14	21	29	38	49		
9	15	22	31	40	51	63	
10	15	23	32	42	53	65	78
12	17	26	35	46	58	71	85
14	19	28	38	50	63	76	91
16	21	31	42	54	67	82	97
18	22	33	45	58	72	87	103
20	24	35	48	62	77	93	110

Теоретические значения W- критерия Вилкоксона
(данные или меньшие значения приводят к опровержению H_0).

Число пар, n_1	Уровни значимости		Число пар, n_1	Уровни значимости	
	0.05	0.01		0.05	0.01
6	1	–	16	31	21
7	3	–	17	36	24
8	5	1	18	41	29
9	7	3	19	47	33
10	9	4	20	53	39
11	12	6	21	60	44
12	15	8	22	67	50
13	18	11	23	74	56
14	22	14	24	82	62
15	26	17	25	90	69

Теоретические значения Z-критерия

Число пар, n_1	Уровни значимости		Число пар, n_1	Уровни значимости	
	0.05	0.01		0.05	0.01
6	6	–	16	13	14
7	7	–	17	13	15
8	8	8	18	14	15
9	8	9	19	15	16
10	9	10	20	15	17
11	10	11	21	16	17
12	10	11	22	17	18
13	11	12	23	17	19
14	12	13	24	18	19
15	12	13	25	18	20

Библиографический список

1. Байбородова, Л. В. Методология и методы научного исследования : Учебное пособие / Л. В. Байбородова, А. П. Чернявская. – 2-е изд., испр. и доп. – Москва : Издательство Юрайт, 2019. – 221 с.
2. Доспехов Б.А. Методика полевого опыта (с основами статистической обработки результатов исследований). Изд-во «АЛЪЯНС», 2011.
3. Калашникова, Е. А. Основы биотехнологии / Е. А. Калашникова, М. Ю. Черденченко, Р. Н. Киракосян. – 2-е издание, исправленное и дополненное. – Москва : ООО "Издательство "КноРус", 2022. – 278 с.
4. Кирюшин, Б. Д. Основы научных исследований в агрономии : учебник для студентов высших учебных заведений, обучающихся по агрономическим специальностям и направлениям / Б. Д. Кирюшин, Р. Р. Усманов, И. П. Васильев. – Москва: КолосС, 2009. – (Учебники и учебные пособия для студентов высших учебных заведений).
5. Колодязная В.С. Методология научных исследований в пищевой биотехнологии /В.С. Колодязная, Е.И. Кипрушкина, Д.А. Бараненко, И.А. Шестопалова, Ю.В. Бройко: Учеб. пособие. – СПб.: Университет ИТМО, 2019. –143 с.
6. Кравченко И.Н. Основы научных исследований / И. Н. Кравченко, А. В. Коломейченко, В. Н. Логачев [и др.]. – 1-е, Новое. – Санкт-Петербург : Издательство Лань, 2015. – 304 с.
7. Кузнецов, И. Н. Основы научных исследований : Учебное пособие / И. Н. Кузнецов. – 4-е издание. – Москва : Издательско-торговая корпорация "Дашков и К", 2018. – 284 с.
8. Усманов Р.Р. Методика экспериментальных исследований в агрономии : учебное пособие для вузов / Р. Р. Усманов.– Москва: Издательство Юрайт, 2021. – 197 с. – (Высшее образование).
9. Усманов Р. Р. Методика опытного дела (с расчетами в программе Excel) : практикум / Р.Р.Усманов, Н.Ф.Хохлов – Москва : РГАУ-МСХА имени К. А. Тимирязева, 2020. – 154 с. – Текст: электронный
10. Шкляр, М. Ф. Основы научных исследований : Учебное пособие для бакалавров / М. Ф. Шкляр. – Москва : Дашков и К, 2016. – 208 с.

Учебное издание

Усманов Раиф Рафикович

**ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ
В БИОТЕХНОЛОГИИ
(с расчетами в программе Excel)**

Учебно-практическое пособие

Ответственный редактор Е.Е. Рытова

Подписано для размещения в Электронно-библиотечной системе
РГАУ-МСХА имени К.А. Тимирязева 2022 г.

Оригинал-макет подготовлен Издательством РГАУ-МСХА
127550, Москва, Тимирязевская ул., 44
Тел. 8 (499) 977-40-64