

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Оренбургский государственный университет»

Кафедра технологии пищевых производств

П.В. Медведев,
В.А. Федотов

МАТЕМАТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

Методические указания

Рекомендовано к изданию редакционно-издательским советом федерального государственного бюджетного образовательного учреждения высшего образования «Оренбургский государственный университет» для обучающихся по образовательной программе высшего образования по направлению подготовки 19.04.02 Продукты питания из растительного сырья

Оренбург
2019

УДК 664.65.05 (075.8)
ББК 36.83-5я73
М 42

Рецензент – доктор технических наук, профессор В. Ю. Полищук

Медведев, П.В.
М 42 Математическая обработка результатов исследования: методические указания / П. В. Медведев, В. А. Федотов; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2019. – 42 с.

Методические указания для выполнения лабораторных работ по дисциплине «Математическая обработка результатов исследования»; способы использования математических методов при обработке, описании, интерпретации результатов исследований.

Методические указания предназначены для обучающихся по образовательной программе высшего образования по направлению подготовки 19.04.02 Продукты питания из растительного сырья.

УДК 664.65.05 (075.8)
ББК 36.82-5я73

© Медведев П.В.,
Федотов В.А., 2019
© ОГУ, 2019

Содержание

Введение	4
1 Основы планирования эксперимента.....	5
2 Лабораторные работы	6
2.1 Лабораторная работа № 1. Критерии нормального распределения	6
2.2 Лабораторная работа № 2. Вывод уравнения простой линейной регрессии	10
2.3 Лабораторная работа № 3. Множественная корреляция в MS Excel	12
2.4 Лабораторная работа № 4. Описательные статистики в среде MS Excel	17
2.5 Лабораторная работа № 5. Расчет коэффициента корреляции Пирсона	24
2.6 Лабораторная работа № 6. Использование при проведении исследований прикладной программы «Полный факторный анализ»	30
Список использованных источников	40

Введение

На основе многолетнего опыта применения методов математического планирования эксперимента при решении различных научных и инженерных задач и анализа работ, изданных в этой области за последние 50 лет, сделана попытка оказания помощи всем, кто связан с постановкой экспериментальных исследований, обработкой результатов опытов и выработкой практических рекомендаций.

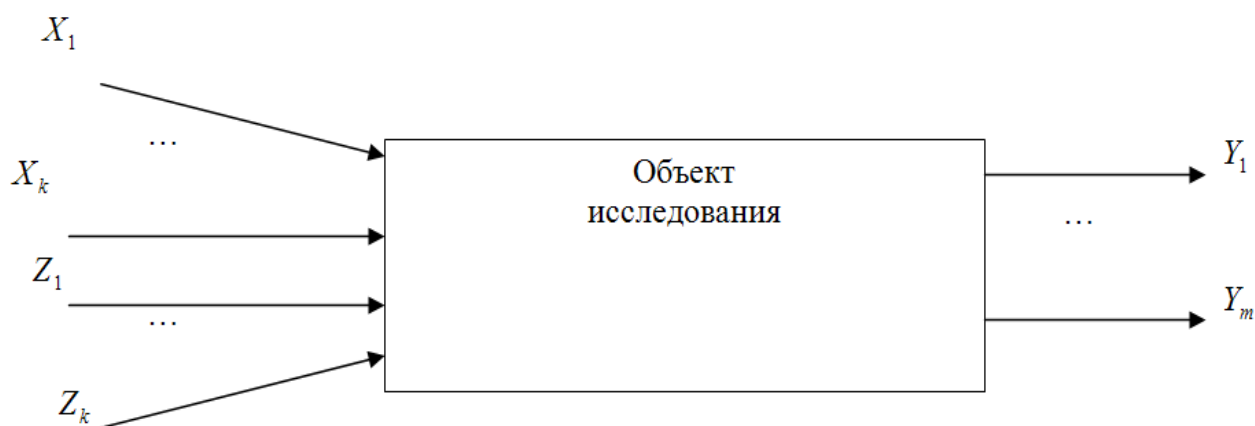
Отличительной особенностью работы является наличие практических примеров из различных областей человеческой деятельности и раскрытие в доступной форме ряда существенных вопросов: выбор плана эксперимента, поиск научной новизны при решении задач, постановка эксперимента в случае невозможности соблюдения требований активного плана, охват области определения многомерной функции, учет в моделях доверительных интервалов и при необходимости качественных и статистически незначимых факторов.

В связи с практической направленностью работы теоретическим вопросам отведено мало внимания, в случае необходимости с ними подробно можно ознакомиться по предлагаемой литературе.

Для научных работников, инженеров, аспирантов, магистрантов и студентов, работающих или обучающихся по различным техническим и другим направлениям.

1 Основы планирования эксперимента

Часто, приступая к изучению какого-либо процесса экспериментатор не имеет исчерпывающих сведений о механизме процесса. Можно только указать параметры определяющие условия протекания процесса, и, возможно требования к его результатам. Поставленная проблема является задачей кибернетики. Действительно, если считать кибернетику «наукой, изучающей системы любой природы, способные воспринимать, хранить и перерабатывать информацию для целей оптимального управления» [1-5], то такую систему можно представить в виде черного ящика. Черный ящик (рисунок 1.1) – объект исследования, имеющий $(k + p)$ входов и m выходов.



X – управляемые, Z – неуправляемые параметры.

Рисунок 1.1 – Модель «черного ящика»

Зависимость между выходными параметрами (откликом) и входными параметрами (факторами) называется функцией отклика. Математическая запись функции отклика представлена в виде формулы

$$y = f(x_1, x_2, \dots, x_k). \quad (1.1)$$

Правильно поставленный эксперимент позволяет проверять гипотезы о причинно-следственных отношениях, не ограничиваясь констатацией связи (корреляции) между переменными.

Среди методов планирования эксперимента различают статические и последовательные. Под статическим планированием эксперимента понимают проведение всего эксперимента по заранее заданному плану. Обычно при статическом планировании используют насыщенные планы.

При последовательном планировании эксперимента происходит перед каждым новым этапом эксперимента. Статическое планирование эксперимента является вычислительно менее затратным, чем последовательное. Это привело к большому развитию статистических методов. Большинство из них рассмотрено в [6-10].

На практике экспериментатору приходится чаще планировать не один, а несколько экспериментов, выполняя и анализируя каждый и, в соответствии с результатами, изменять план эксперимента.

Использование теории планирования эксперимента является одним из путей существенного повышения эффективности многофакторных экспериментальных исследований. Под планированием эксперимента понимают процедуру выбора числа и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью.

2 Лабораторные работы

2.1 Лабораторная работа № 1. Критерии нормального распределения

Встречаемые в обыденной жизни большое число наблюдений, измеряемые значения параметров, погрешности измерений и ошибки подчиняются нормальному закону. Этот закон можно считать основой математической статистики.

Нормальным называется закон распределения случайной величины x , если плотность распределения вероятности определяется по формуле

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\bar{x})^2}, \quad (-\infty < x < +\infty, \sigma > 0), \quad (2.1)$$

где σ – среднее квадратическое отклонение,

\bar{x} - среднее значение (математическое ожидание),

$\pi = 3,141593$, $e = 2,718282$ – математические константы.

Из формулы следует, что нормальное распределение полностью определяется параметрами \bar{x} и σ . Среднее квадратическое отклонение определяет форму кривой: чем больше σ (разброс данных), тем кривая становится пологой. Математическое ожидание \bar{x} определяет положение кривой на оси абсцисс, кривая симметрична относительно этого значения.

Нормальное распределение часто встречается в природе. Например, следующие случайные величины хорошо моделируются нормальным распределением:

- отклонение при стрельбе;
- погрешности измерений (однако погрешности некоторых измерительных приборов имеют не нормальные распределения);
- рост человека;
- давление крови в течение дня;
- экзаменационные оценки;
- некоторые характеристики живых организмов в популяции.

Многие непрерывные случайные величины не являются ни точно, ни приближенно нормальными. Свойства таких величин довольно сильно отличаются от свойств нормального распределения, перечисленных выше.

Вероятность, что подброшенная монета упадет орлом вверх 50 %, что при броске шестигранного кубика выпадет сторона «4» - 16,7 %, что завтра на когонибудь упадет метеорит – 0,00000000294 %. Это простые примеры, достаточно

разделить количество желаемых событий на общее количество случаев и мы получаем вероятность события, но когда результаты эксперимента могут быть не только орлом или решкой (что эквивалентно да/нет), а большим набором данных. Например, вес батона хлеба, если мы возьмём в магазине 1000 буханок хлеба и взвесим каждую, то мы узнаем, что на самом деле батон не весит 400 грамм, результаты будут варьироваться в диапазоне от 384 до 416 грамм (допуск разброса веса предусмотрен ГОСТ).

Такую форму график получит потому, что большинство значений близко к 400. Это - пример нормального распределения, множество событий имеют закон нормального распределения, например, вес или рост для определённого возраста, или среднее время Вашего похода до магазина и многие другие события также подчиняются закону нормального распределения. Так работают маркетологи: проводят опрос 1000 человек и получают представление о всём населении.

В случае таблицы Вы имеете дело с дискретными данными, т.е. для каждого веса есть определённая вероятность, но в случае графика дело немного меняется, теперь мы говорим не о 1000 буханок, которые мы взвесили, а обо всех буханках в мире сразу, чтобы не взвешивать все буханки. Имея закон распределения, который мы получили взвесив 1000 буханок (мы могли взвесить 100, 200, 500, сколько угодно), мы можем предположить, что сколько бы мы буханок не взяли, замерив их, мы получим ту же форму колокола. Используя термины статистики, все буханки хлеба - это генеральная совокупность, 1000 замеренных буханок - выборка.

Нормальное распределение получило своё название абсолютно справедливо: по статистике, большинство событий происходят именно с вероятностью нормального распределения, но что это значит? Это означает, например, что когда Вы видите на упаковке хлеба обозначение «Вес: 400 ± 16 г» - вес батона имеет нормальное распределение со средним значением 400 г и стандартным отклонением 16 г.

Задание: использовать для исследований не менее 10 образцов одного вида хлеба для получения сведений о распределении случайной величины (веса буханки). Отобразить в отчете графическое представление данных о плотности вероятности

нормального распределения. Проверку нормальности распределения случайной величины можно производить разными способами, рассмотрим простую рекомендацию. Надо вычислить среднее абсолютное по обычной формуле

$$e = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}. \quad (2.2)$$

Если выполняется условие

$$\left| \frac{e}{S} - 0,7979 \right| < \frac{0,4}{\sqrt{n}}, \quad (2.3)$$

то гипотеза нормальности распределения выборки данных принимается. В противном случае надо постараться преобразовать распределение к нормальному или применить другое распределение.

Задание: занести в таблицу 2.1 данные анализа показателей качества зерна имеющихся образцов пшеницы (желательно более 5). Построить диаграммы распределения этих показателей качества и оценить «нормальность» распределения данных.

Таблица 2.1 – Данные анализа показателей качества зерна

№ образца	Число падения, с	Количество клейковины, %	Качество клейковины, ед. ИДК

1.13.1 Проверить на нормальное распределение ряд чисел 21, 89, 54, 25, 9, 10, 14, 47, 74, 21, 89, 54, 25, 9, 1, 47, 97.

1.13.2 Установить, соответствие критериям нормальности численного ряда 89, 54, 25, 9, 1, 47, 97, 74, 9, 10, 14, 32, 47, 20, 15, 14.

1.13.3 Построить диаграмму нормальности для ряда чисел 97, 74, 9, 10, 14, 32, 47, 54, 78, 12, 45, 56, 54.

2.2 Лабораторная работа № 2. Вывод уравнения простой линейной регрессии

2.2.1 Постановка задачи

Вернемся к сценарию, изложенному в начале. Наша цель - предсказать объем годовых продаж для всех новых магазинов, зная их размеры. Для оценки зависимости между размером магазина (в квадратных футах) и объемом его годовых продаж создадим выборки из 14 магазинов.

Анализ рисунка показывает, что между площадью магазина X и годовым объемом продаж Y существует положительная зависимость. Если площадь магазина увеличивается, объем продаж возрастает почти линейно. Таким образом, наиболее подходящей для исследования является линейная модель. Остается лишь определить, какая из линейных моделей точнее остальных описывает зависимость между анализируемыми переменными. [19]

2.2.2 Метод наименьших квадратов

Данные получены для случайной выборки магазинов. Если верны некоторые предположения, в качестве оценки параметров генеральной совокупности (β_0 и β_1) можно использовать сдвиг b_0 и наклон b_1 прямой Y . Таким образом, уравнение простой линейной регрессии принимает следующий вид

$$Y_i = b_0 + b_1 \cdot X_i, \quad (2.4)$$

где Y_i - предсказанное значение переменной Y для i -того наблюдения,

X_i - значение переменной X в i -том наблюдении.

Для того чтобы предсказать значение переменной Y , в уравнении необходимо определить два коэффициента регрессии – сдвиг b_0 и наклон b_1 прямой Y . Вычислив эти параметры, проведем прямую на диаграмме разброса. Затем исследователь может визуальнo оценить, насколько близка регрессионная прямая к точкам наблюдения. Простая линейная регрессия позволяет найти прямую линию, максимально приближенную к точкам наблюдения. Критерии соответствия можно задать разными способами. Возможно, проще всего минимизировать разности между фактическими значениями Y_i и предсказанными значениями \hat{Y}_i . Однако, поскольку эти разности могут быть как положительными, так и отрицательными, следует минимизировать сумму их квадратов. Сумма квадратов принимает следующий вид

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - (b_0 + b_1 X_i))^2 \quad (2.5)$$

Параметры b_0 и b_1 неизвестны. Таким образом, сумма квадратов разностей является функцией, зависящей от сдвига b_0 и наклона b_1 выборки Y . Для того чтобы найти значения параметров b_0 и b_1 , минимизирующих сумму квадратов разностей, применяется метод наименьших квадратов. При любых других значениях сдвига b_0 и наклона b_1 сумма квадратов разностей между фактическими значениями переменной Y и ее наблюдаемыми значениями лишь увеличится.

До того, как Excel взял на себя всю рутинную работу, вычисления по методу наименьших квадратов были очень трудоемкими. Excel позволяет решать подобные задачи двумя способами. Во-первых, можно воспользоваться Пакетом анализа (строка Регрессия).

Во-вторых, можно, выделив точки на графике, кликнуть правой кнопкой мыши и выбрать Добавить линию тренда. Далее можно выбрать вид линии тренда (в нашем случае - Линейная), отформатировать линию, показать на графике уравнение и величину достоверности аппроксимации (R^2). Диаграмма разброса и линия регрессии (тренда) в задаче о выборе магазина - $b_0 = 0,9645$, а $b_1 = 1,6699$. Таким образом, уравнение линейной регрессии для этих данных имеет следующий вид [20]

$$Y_i = 0,9645 + 1,6699 \cdot X_i. \quad (2.6)$$

2.2.3 Задание для проведения регрессионного анализа в MS Excel

В таблице 2.2 представлены данные анализа показателей качества макаронных изделий более 40 образцов пшеницы. Требуется оценить характер зависимости между отдельными показателями качества макарон.

2.3 Лабораторная работа № 3. Множественная корреляция в MS Excel

2.3.1 Возможности MS Excel для проведения корреляционного анализа

При большом числе наблюдений, когда коэффициенты корреляции необходимо последовательно вычислять для нескольких выборок, для удобства получаемые коэффициенты сводят в таблицы, называемые корреляционными матрицами. Корреляционная матрица - это квадратная таблица, в которой на пересечении соответствующих строк и столбцов находятся коэффициент корреляции между соответствующими параметрами.

Таблица 2.2 - Показатели качества макаронных изделий из зерна пшеницы

Образец муки №	Показатели качества макаронных изделий			
	Коэффициент увеличения массы	Сухие вещества, перешедшие в воду при варке, %	Время варки до готовности, мин	Прочность сухих изделий на срез, Н
1	1,4	4,3	10,8	4,9
2	1,6	4,7	11,2	5,4
3	1,7	4,1	9,7	5,9
4	1,6	4,8	12,3	5,8
5	1,5	4,8	12,2	5,9
6	1,8	4,4	11,8	3,2
7	1,2	5,6	11,3	5,8
8	1,5	5,4	13	6,6
9	1,8	5,7	13,2	6,6
10	1,8	5,8	10,1	6,7
11	1,9	5,7	10,7	5,4
12	1,9	5	13,6	5,3
13	1,6	5	12,5	7,9
14	2,1	5,8	10,3	6,7
15	1,6	4,9	12,6	5,6
16	1,8	6	10,9	4,6
17	2,1	5,6	11,8	7,9
18	2	6,3	13,3	7,6
19	2,1	5,9	13	7,6
20	1,9	5,9	14,3	6,8
21	2,1	6	13,6	6,6
22	1,8	5,4	12,8	8,7

Продолжение таблицы 2.2

Образец муки №	Показатели качества макаронных изделий			
	Коэффициент увеличения массы	Сухие вещества, перешедшие в воду при варке, %	Время варки до готовности, мин	Прочность сухих изделий на срез, Н
23	2,3	6,1	13,9	8,2
24	2,3	5,9	12,7	7
25	1,9	5,3	13	8,5
26	2,1	5,3	10,4	6,1
27	1,8	6,3	13,7	6
28	2,1	5,7	14,2	6,9
29	2,4	6	13,5	7,6
30	1,9	5,2	11,9	8,1
31	2,1	3,1	13	6,3
32	2,3	5,8	14,4	8,9
33	1,7	6,4	12,3	7,6
34	2,3	5,7	15,3	7
35	2,4	6	15,2	7,8
36	2,3	5,6	14,1	8,4
37	2,1	7	11,7	8,5
38	2	6,2	16,1	8,9
39	1,7	6,8	15	5,4
40	2,3	5,7	12,7	7
41	2,4	5,1	14,0	6,9

В MS Excel для вычисления корреляционных матриц используется процедура Корреляции из пакета Анализ данных. Процедура позволяет получить корреляционную матрицу, содержащую коэффициенты корреляции между различными параметрами.

Для реализации процедуры необходимо:

1. Выполнить команду Сервис - Анализ данных;
2. В появившемся списке Инструменты анализа выбрать строку Корреляция и нажать кнопку ОК;
3. В появившемся диалоговом окне указать Входной интервал, то есть ввести ссылку на ячейки, содержащие анализируемые данные. Входной интервал должен содержать не менее двух столбцов.
4. В разделе Группировка переключатель установить в соответствии с введенными данными (по столбцам или по строкам);
5. Указать выходной интервал, то есть ввести ссылку на ячейку, начиная с которой будут показаны результаты анализа. Размер выходного диапазона будет определен автоматически, и на экран будет выведено сообщение в случае возможного наложения выходного диапазона на исходные данные. Нажать кнопку ОК.

В выходной диапазон будет выведена корреляционная матрица, в которой на пересечении каждой строки и столбца находится коэффициент корреляции между соответствующими параметрами. Ячейки выходного диапазона, имеющие совпадающие координаты строк и столбцов, содержат значение 1, так как каждый столбец во входном диапазоне полностью коррелирует сам с собой.

Разберем алгоритм проведения корреляционного анализа в среде MS Excel на примере анализа зависимости урожайности зерновых культур от качества земли.

Имеются данные по 20 сельскохозяйственным хозяйствам. Найти коэффициент корреляции между величинами урожайности зерновых культур и качеством земли и оценить его значимость. Данные приведены в таблице 2.3.

Таблица 2.3 - Зависимость урожайности зерновых культур от качества земли

Номер хозяйства	Качество земли, балл	Урожайность, ц/га
1	32	19,5
2	33	19
3	35	20,5
4	37	21
5	38	20,8
6	39	21,4
7	40	23
8	41	23,3
9	42	24
10	44	24,5
11	45	24,2
12	46	25
13	47	27
14	49	26,8
15	50	27,2
16	52	28
17	54	30
18	55	30,2
19	58	32
20	60	33

Решение. Для выполнения корреляционного анализа введите в диапазон A1:G3 исходные данные. Затем в меню Сервис выберите пункт Анализ данных и далее укажите строку Корреляция. В появившемся диалоговом окне укажите Входной интервал (A2:C7). Укажите, что данные рассматриваются по столбцам. Укажите выходной диапазон (E1) и нажмите кнопку ОК.

Существует другой способ проведения такого анализа без вызова промежуточных меню и окон. Ставим курсор в любую ячейку и нажимаем кнопку «fx».

В категории «Статистические» выбираем функцию КОРРЕЛ. Аргумент «Массив 1» - первый диапазон значений – время работы станка: A2:A14. Аргумент «Массив 2» - второй диапазон значений – стоимость ремонта: B2:B14. Жмем ОК.

Чтобы определить тип связи, нужно посмотреть абсолютное число коэффициента (для каждой сферы деятельности есть своя шкала). Для корреляционного анализа нескольких параметров (более 2) удобнее применять «Анализ данных» (надстройка «Пакет анализа»). В списке нужно выбрать корреляцию и обозначить массив. Полученные коэффициенты отобразятся в корреляционной матрице.

2.3.2 Задание для проведения корреляционного анализа в MS Excel

В таблице 2.4 представлены данные анализа показателей качества зерна пшеницы. Требуется определить корреляционные зависимости между отдельными показателями качества зерна.

2.4 Лабораторная работа № 4. Описательные статистики в среде MS Excel

2.4.1 Основные статистические параметры выборки

Рассмотрим инструмент Описательная статистика, входящий в надстройку Пакет Анализа. Рассчитаем показатели выборки: среднее, медиана, мода, дисперсия, стандартное отклонение и др.

Таблица 2.4 – Показатели качества зерна пшеницы Оренбургской области

Образец муки №	Показатели качества зерна				
	Кислотность, град	Влажность, %	Количество клейковины, %	Качество клейковины, единиц ИДК	Твердозерность, кг/мм ²
1	2	3	4	5	6
1	2,2	12,3	21,2	81	12,2
2	2,1	14,2	20,9	66	15,4
3	2,1	15,9	23,7	51	17,6
4	2,5	14,9	26,4	73	19,1
5	3,1	15,1	26	46	21,5
6	3,0	12,9	28,1	47	23,8
7	1,5	11,1	28,4	38	25,6
8	1,7	13,6	21,5	75	13,5
9	1,9	15,2	22,2	89	14,8
10	2,8	11,7	22,6	81	16,1
11	2,4	14,3	24	56	18,0
12	3,8	13,8	27,3	44	20,7
13	1,7	14,3	22,2	82	15,5
14	1,9	13	25,2	79	15,7
15	2,7	14,3	25,8	73	15,9
16	1,9	15,2	25,1	89	16,2
17	3,4	12	24,6	58	16,5
18	3,2	15,4	24,7	75	17,0
19	1,8	13,1	24,8	87	17,4
20	3,4	13,8	22,7	87	17,8
21	2,5	11,7	22,1	50	18,2

Продолжение таблицы 2.4

1	2	3	4	5	6
22	2,6	15,2	23,1	73	18,5
23	3,5	12,3	27,2	47	18,9
24	1,4	11,9	26,2	57	19,1
25	1,9	14,5	22,9	54	19,3
26	1,4	12,1	23	55	19,5
27	3,2	14,7	27,7	43	20,0
28	3,4	13,1	27,9	62	20,2
29	2,7	14,3	25,5	75	20,5
30	2,9	14,1	23,8	41	20,8
31	3,4	11,9	24,9	76	21,0
32	3,5	13,4	28	38	21,1
33	2,6	15,5	26,2	50	21,3
34	2,1	15,3	25,4	72	21,4
35	2,8	13,3	25,8	56	21,6
36	3,1	13,4	25	37	21,9
37	2,3	11,7	24,3	66	22,3
38	2,5	13,4	28,7	66	22,5
39	2,3	13,7	26,5	52	22,7

Задача описательной статистики (descriptive statistics) заключается в том, чтобы с использованием математических инструментов свести сотни значений выборки к нескольким итоговым показателям, которые дают представление о выборке.

В диалоговом окне Анализ данных выберите инструмент Описательная статистика. После нажатия кнопки ОК будет выведено другое диалоговое окно, в котором нужно указать:

- входной интервал (Input Range) – это диапазон ячеек, в котором содержится массив данных. Если в указанный диапазон входит текстовый заголовок набора

данных, то нужно поставить галочку в поле Метки в первой строке (Labelsinfirstrow). В этом случае заголовок будет выведен в Выходном интервале. Пустые ячейки будут проигнорированы, поэтому нулевые значения необходимо обязательно указывать в ячейках, а не оставлять их пустыми;

- выходной интервал (Output Range). Здесь укажите адрес верхней левой ячейки диапазона, в который будут выведены статистические показатели;

- итоговая статистика (SummaryStatistics). Поставьте галочку напротив этого поля – будут выведены основные показатели выборки: среднее, медиана, мода, стандартное отклонение и др. [21]

Также можно поставить галочки напротив полей Уровень надежности (ConfidenceLevelforMean), К-й наименьший (Kth Largest) и К-й наибольший (Kth Smallest). В результате будут выведены следующие статистические показатели.

Все показатели выведены в виде значений, а не формул. Если массив данных изменился, то необходимо перезапустить расчет. Если во входном интервале указать ссылку на несколько столбцов данных, то будет рассчитано соответствующее количество наборов показателей. Такой подход позволяет сравнить несколько наборов данных. При сравнении нескольких наборов данных используйте заголовки (включите их во Входной интервал и установите галочку в поле Метки в первой строке). Если наборы данных разной длины, то это не проблема - пустые ячейки будут проигнорированы.

Зеленым цветом на картинке выше и в файле примера выделены показатели, которые не требуют особого пояснения. Для большинства из них имеется специализированная функция.

Интервал (Range) - разница между максимальным и минимальным значениями.

Минимум (Minimum) – минимальное значение в диапазоне ячеек, указанном во Входном интервале.

Максимум (Maximum) – максимальное значение.

Сумма (Sum) – сумма всех значений.

2.4.2 Задание для определения основных статистических параметров выборки в MS Excel

В таблицах 2.5 и 2.6 представлены данные анализа показателей качества зерна пшеницы. Необходимо найти основные статистические характеристики образцов зерна по годам урожая и по зонам произрастания, сравнить их, сделать выводы о преобладании влияния погодных или географических условий произрастания.

Таблица 2.5 – Данные урожайности зерна за 5 лет (в ц/га)

Сорт	Зона произрастания / Год урожая	2012	2013	2014	2015	2016
		1	2	3	4	5
Харьковская 3	восточная	15,3	15,5	16,1	16,8	14,7
	центральная	15	16,7	15,2	17,4	17,8
	западная	16,5	17,5	16,5	16,7	16,8
Оренбургская 10	восточная	16,1	17,4	16,6	17,6	16,3
	центральная	15,9	18,7	17,9	17,5	18,7
	западная	15,7	18,4	18,1	16,7	15,3
Оренбургская 21	восточная	17,2	17,4	17,3	16,9	18,4
	центральная	16,9	18,1	18,8	19,3	19
	западная	19	16,4	18,3	17,2	17,5
Безенчукская Янтарь	восточная	17,1	17,9	15,5	15,8	18,6
	центральная	17,9	19,3	18,7	17,7	17,1
	западная	18,9	18,1	16,4	18,6	19,3
Безенчукская 200	восточная	18	15,7	20	19,3	18,9
	центральная	18,1	19,3	19	19,3	19,3
	западная	17	17	17,1	19,2	19,6

Продолжение таблицы 2.5

1	2	3	4	5	6	7
Степь 3	восточная	15,8	20,5	16,3	16,8	18
	центральная	18,4	19,8	17,8	17,5	19,2
	западная	18,8	20,4	19,3	18,3	17,2
Юго- Восточная 3	восточная	17,4	16,6	16,8	16,9	21,8
	центральная	20,8	21,7	20,8	20,9	16,7
	западная	19,7	21,3	19,6	20	18,5
Учитель	восточная	21,5	19,5	19,6	20,7	13,4
	центральная	15,9	19,1	17,8	16,8	16,2
	западная	17,5	21,8	17,6	21,7	17,7
Варяг	восточная	17,4	19,3	19,5	17,1	20,7
	центральная	17,9	21,7	21,2	20,1	21,6
	западная	19,1	19,8	17,8	21,7	16,8
Оренбургская 13	восточная	18,4	21,3	17,4	17,3	17,7
	центральная	20,8	19,6	22,2	19,3	20,1
	западная	18,3	17,6	21,4	22,8	16,5
Прохор	восточная	21,2	20,9	14,3	22,3	20,3
	центральная	21,5	20,7	21,1	19,8	18,5
	западная	14,9	22,6	18,5	19,1	19
Л 503	восточная	16,7	17,7	21,3	16,7	15,8
	центральная	22,2	20,7	20,7	21,8	19,3
	западная	18,9	21,7	21,4	20	20,9
Саратовская 42	восточная	17,9	20,2	24,7	22,1	18,8
	центральная	17,1	20,8	19,5	22,6	21,4
	западная	19,4	15,1	20,6	20,7	15,6

Таблица 2.6 – Показатель седиментации зерна пшеницы за 5 лет (в см³)

Сорт	Зона произрастания / Год урожая	2012	2013	2014	2015	2016
		1	2	3	4	5
Харьковская 3	восточная	12	18	10	10	21
	центральная	10	11	19	14	24
	западная	10	14	25	16	25
Оренбургская 10	восточная	16	15	22	18	18
	центральная	17	11	14	21	27
	западная	20	16	18	20	29
Оренбургская 21	восточная	17	18	13	25	26
	центральная	18	17	12	19	17
	западная	16	26	27	22	28
Безенчукская Янтарь	восточная	18	21	24	26	30
	центральная	17	27	20	26	33
	западная	30	14	17	17	10
Безенчукская 200	восточная	26	20	23	19	21
	центральная	19	24	17	30	25
	западная	17	27	14	14	21
Степь 3	восточная	22	14	20	15	24
	центральная	30	23	16	29	25
	западная	20	25	19	26	18
Юго- Восточная 3	восточная	37	16	36	18	14
	центральная	24	14	23	29	32
	западная	24	11	32	12	33
Учитель	восточная	34	20	27	26	30
	центральная	25	36	28	19	21
	западная	41	40	34	30	20

Продолжение таблицы 2.6

1	2	3	4	5	6	7
Варяг	восточная	18	11	10	38	19
	центральная	35	26	39	31	33
	западная	36	14	46	21	39
Оренбургская 13	восточная	19	27	30	25	33
	центральная	30	22	39	33	38
	западная	24	34	16	41	28
Прохор	восточная	26	48	30	20	16
	центральная	45	15	39	41	14
	западная	27	59	40	38	32
Л 503	восточная	28	20	15	21	33
	центральная	55	31	25	33	30
	западная	18	51	55	18	21
Саратовская 42	восточная	37	32	55	22	20
	центральная	24	21	36	43	19
	западная	24	15	23	28	33

2.5 Лабораторная работа № 5. Расчет коэффициента корреляции Пирсона

Одним из критериев, позволяющих определить, какова теснота (или сила) корреляционной связи между двумя показателями, измеренными в количественной шкале является критерий корреляции Пирсона.

Целью исследования явилось выявление, определение тесноты и статистической значимости корреляционной связи между двумя количественными показателями: газообразующей способностью муки ГОСМ (X) и показателя

амилолитической активности муки («число падения» или ЧП) - (Y). Исходные данные для выборки, состоящей из 5 опытов (n = 5), сведены в таблице 2.7.

Таблица 2.7 – Исходные данные по результатам экспериментов

№	Газообразующая способность муки, мл CO ₂ на 100 г (X)	Число падения, с (Y)
1	951	83
2	874	76
3	957	84
4	1084	89
5	903	79

Вычислим суммы анализируемых значений X и Y

$$\Sigma(X) = 951 + 874 + 957 + 1084 + 903 = 4769,$$

$$\Sigma(Y) = 83 + 76 + 84 + 89 + 79 = 441.$$

Найдем средние арифметические для X и Y

$$M_x = \Sigma(X) / n = 4769 / 5 = 953,8,$$

$$M_y = \Sigma(Y) / n = 441 / 5 = 88,2.$$

Рассчитаем для каждого значения сопоставляемых показателей величину отклонения от среднего арифметического d_x и d_y (таблица 2.8).

Таблица 2.8 – Расчет отклонений значений экспериментов от среднего

№	Газообразующая способность муки, мл CO ₂ на 100 г (X)	Число падения, с (Y)	Отклонение ГОСМ от среднего значения (d_x)	Отклонение % ЧП от среднего значения (d_y)
1	951	83	-2,8	0,8
2	874	76	-79,8	-6,2
3	957	84	3,2	1,8
4	1084	89	130,2	6,8
5	903	79	-50,8	-3,2

Возведем в квадрат каждое значение отклонения d_x и d_y – таблица 2.9.

Рассчитаем для каждой пары анализируемых значений произведение отклонений d_x и d_y (таблица 2.10). Определим значения суммы квадратов отклонений $\sum d_x^2$ и $\sum d_y^2$

$$\sum d_x^2 = 25918,8$$

$$\sum d_y^2 = 98,8$$

Найдем значение суммы произведений отклонений $\sum (d_x \cdot d_y)$

$$\sum (d_x \cdot d_y) = 1546,2$$

Таблица 2.9 – Расчеты промежуточных статистических данных

№	Газообразующая способность муки, мл CO ₂ на 100 г (X)	Число падений, с (Y)	Отклонение ГОСМ от среднего значения (d_x)	Отклонение % ЧП от среднего значения (d_y)	d_x^2	d_y^2
1	951	83	-2,8	0,8	7,84	0,64
2	874	76	-79,8	-6,2	6368,04	38,44
3	957	84	3,2	1,8	10,24	3,24
4	1084	89	130,2	6,8	16952,04	46,24
5	903	79	-50,8	-3,2	2580,64	10,24

Рассчитаем значение коэффициента корреляции Пирсона r по формуле

$$r = \frac{\sum(d_x \cdot d_y)}{\sqrt{(\sum d_x^2 \cdot \sum d_y^2)}}, \quad (2.7)$$

$$r = \frac{1546,2}{\sqrt{25918,8 \cdot 98,8}} = 0,966$$

Коэффициент корреляции изменяется от -1 до +1. Если значение положительное то это свидетельствует о прямой связи, отрицательное значение – об обратной. Характеристики тесноты связи отображены в таблицах 2.11 и 2.12.

Таблица 2.10 – Значения промежуточных статистических данных

№	Газообразующая способность муки, мл CO ₂ на 100 г (X)	Число падений, с (Y)	Отклонение ГОСМ от среднего значения (d_x)	Отклонение % ЧП от среднего значения (d_y)	d_x^2	d_y^2	$d_x \cdot d_y$
1	951	83	-2,8	0,8	7,84	0,64	-2,24
2	874	76	-79,8	-6,2	6368,04	38,44	494,76
3	957	84	3,2	1,8	10,24	3,24	5,76
4	1084	89	130,2	6,8	16952,04	46,24	885,36
5	903	79	-50,8	-3,2	2580,64	10,24	162,56

Таблица 2.11 – Таблица Чеддока для установления силы связи

Абсолютное значение r	Теснота (сила) корреляционной связи
менее 0,3	слабая
от 0,3 до 0,5	умеренная
от 0,5 до 0,7	заметная
от 0,7 до 0,9	высокая
более 0,9	весьма высокая

Найдем значение t-критерия для оценки статистической значимости корреляционной связи

$$t_r = \frac{0,97 \cdot \sqrt{5-2}}{\sqrt{1-0,97^2}} = 7,0$$

Критическое значение t-критерия найдем по таблице, где при числе степеней свободы $f = n - 2 = 3$ и уровне значимости $p = 0,01$ значение $t_{table} = 5,84$. Рассчитанное значение $t_r (7,0)$ больше $t_{table} (5,84)$, следовательно связь является статистически значимой.

Сделаем статистический вывод. Значение коэффициента корреляции Пирсона составило 0,97, что соответствует весьма высокой тесноте связи между газообразующей способностью муки и амилолитической активностью муки. Данная корреляционная связь является статистически значимой ($p < 0,01$).

Таблица 2.12 – Интерпретация значений линейного коэффициента корреляции

Значения линейного коэффициента корреляции	Характер связи
$r = -1$	Функциональная
$-1 < r < -0,7$	Обратная сильная
$-0,7 \leq r \leq -0,5$	Обратная умеренная
$-0,5 < r < 0$	Обратная слабая
$r = 0$	Отсутствует
$0 < r < +0,5$	Прямая слабая
$+0,5 \leq r \leq +0,7$	Прямая умеренная
$+0,7 < r < +1$	Прямая сильная
$r = +1$	Функциональная

2.6 Лабораторная работа № 6. Использование при проведении исследований прикладной программы «Полный факторный анализ»

Прикладная программа «Полный факторный анализ планов первого порядка» предназначена для быстрого составления планов эксперимента первого порядка и проведения факторного анализа на его основе.

Прикладная программа «Полный факторный анализ планов первого порядка» направлена на составление планов эксперимента первого порядка и проведение факторного анализа на его основе.

Исходные данные могут вбиваться самостоятельно пользователем программы.

Для использования в учебно-практических целях в программе предусмотрено наличие:

- возможности демонстрации полного хода проведения полного факторного эксперимента;
- высокая скорость расчета.

Прикладная программа «Полный факторный анализ планов первого порядка» предназначена для работы в графических операционных средах типа Microsoft Windows, Android, FreeBSD, OpenBSD, Linux, etc.

Проведенные эксперименты показывают, что для работы с данной прикладной программой не требуется специальной подготовки, она осваивается пользователем в течение 10 минут с начала работы с данной программой.

Прикладная программа «Полный факторный анализ планов первого порядка» не требует установки и дополнительной отладки. Для запуска программы необходимо зайти на веб-сайт <http://bread.smartbunny.ru/experiments/>

Исходными данными для программы являются числовые и строчные значения переменных, выбираемые и вводимые пользователем, на основе которых производятся все необходимые расчеты .

Математическая обработка данных, полученных в ходе исследования – важное условие при выполнении курсовых и дипломных работ.

Для анализа многомерных данных часто используется факторный анализ. Под многомерным количественным представлением объекта понимается присвоение ему одновременно несколько численных значений, характеризующих степень выраженности различных его свойств.

Факторным анализом называется любая процедура переработки многомерных данных, на «вход» которой «подается» матрица корреляции между исходными переменными (показателями, шкалами и пр.), а с «выхода» считывается система переменных, именуемых факторами.

Лучше всего заранее, еще на стадии планирования, подбирать математическую модель, которая в дальнейшем будет использоваться или принимать оптимальное решение о выборе адекватной модели с учетом условий проведения исследования и разнообразия существующих математических методов.

Зависимость между выходными параметрами (откликом) и входными параметрами (факторами) называется функцией отклика.

Использование теории планирования эксперимента является одним из путей существенного повышения эффективности многофакторных экспериментальных исследований. В планировании экспериментов применяются в основном планы первого и второго порядков. Планы более высоких порядков используются в инженерной практике редко. В связи с этим далее приводится краткое изложение методики составления планов эксперимента для моделей первого и второго порядков. Под планом первого порядка понимают такие планы, которые позволяют провести эксперимент для отыскания уравнения регрессии, содержащего только первые степени факторов и их произведения.

Первый этап планирования эксперимента для получения линейной модели основан на варьировании факторов на двух уровнях. В этом случае, если число факторов известно, можно сразу найти число опытов, необходимое для реализации всех возможных сочетаний уровней факторов. Простая формула, которая для этого используется

$$N = 2^k, \quad (2.8)$$

где N – число опытов, k – число факторов, 2 – число уровней.

В общем случае эксперимент, в котором реализуются всевозможные сочетания уровней факторов, называется полным факторным экспериментом. Если число уровней каждого фактора равно двум, то имеем полный факторный эксперимент типа 2^k .

Нетрудно написать все сочетания уровней в эксперименте с двумя факторами. Напомним, что в планировании эксперимента используются кодированные значения факторов: $+1$ и -1 (часто для простоты записи единицы опускают). Условия эксперимента можно записать в виде таблицы, где строки соответствуют различным опытам, а столбцы – значениям факторов. Будем называть такие таблицы матрицами планирования эксперимента. Матрица планирования для двух факторов приведена в таблице 2.13.

Таблица 2.13 – Матрица планирования эксперимента при 2 факторах

№ опыта	x_1	x_2	y
1	-1	-1	y_1
2	+1	-1	y_2
3	-1	+1	y_3
4	+1	+1	y_4

Каждый столбец в матрице планирования называют вектор-столбцом, а каждую строку – вектор-строкой. Таким образом, мы имеем 2 вектор-столбца независимых переменных и один вектор-столбец параметра оптимизации.

Если для двух факторов все возможные комбинации уровней легко найти прямым перебором (или просто запомнить), то с ростом числа факторов возникает необходимость в некотором приеме построения матриц. Из многих возможных обычно используется три приема, основанные на переходе от матриц меньшей

размерности к матрицам большей размерности. Рассмотрим первый. При добавлении нового фактора каждая комбинация уровней исходного плана встречается дважды: в сочетании с нижним и верхним уровнями нового фактора. Отсюда естественно появляется прием: записать исходный план для одного уровня нового фактора, а затем повторить его для другого уровня. Вот как это выглядит при переходе от эксперимента 2^2 к 2^3 – таблица 2.14.

Этот прием распространяется на построение матриц любой размерности.

Таблица 2.14 – Матрица планирования эксперимента при 3 факторах

№ опыта	x_1	x_2	x_3	y
1	–	–	+	y_1
2	+	–	+	y_2
3	–	+	+	y_3
4	+	+	+	y_4
5	–	–	–	y_5
6	+	–	–	y_6
7	–	+	–	y_7
8	+	+	–	y_8

Рассмотрим второй прием. Для этого введем правило перемножения столбцов матрицы. При построчном перемножении двух столбцов матрицы произведение единиц с одноименными знаками дает +1, а с разноименными –1. Воспользовавшись этим правилом, получим для случая, который мы рассматриваем, вектор-столбец произведения x_1x_2 в исходном плане. Далее повторим еще раз исходный план, а у столбца произведений знаки поменяем на обратные. Этот прием тоже можно перенести на построение матриц любой размерности, однако он сложнее, тем первый.

Третий прием основан на правиле чередования знаков. В первом столбце знаки меняются поочередно, во втором столбце они чередуются через два, в третьем – через 4, в четвертом – через 8 и т. д. по степеням двойки.

Оценки коэффициентов модели должны быть наилучшими и что точность предсказания параметра оптимизации не должна зависеть от направления в факторном пространстве, ибо заранее неясно, куда предстоит двигаться в поисках оптимума.

Два свойства следуют непосредственно из построения матрицы. Первое из них – симметричность относительно центра эксперимента – формулируется следующим образом: алгебраическая сумма элементов вектор-столбца каждого фактора равна нулю, или, где j – номер фактора, N – число опытов, $i = 1, 2, \dots, k$.

Второе свойство – так называемое условие нормировки – формулируется следующим образом: сумма квадратов элементов каждого столбца равна числу опытов, или

$$\sum_{i=1}^N x_{ji}^2 = N \quad (2.9)$$

Это следствие того, что значения факторов в матрице задаются +1 и –1.

Это свойства отдельных столбцов матрицы планирования. Теперь остановимся на свойстве совокупности столбцов. Сумма почленных произведений любых двух вектор-столбцов матрицы равна нулю, или

$$\sum_{i=1}^N x_{ji}x_{wi} = 0, \quad j \neq w, \quad j, w = 0, 1, 2, \dots, k \quad (2.10)$$

Это важное свойство называется ортогональностью матрицы планирования.

Последнее, четвертое свойство называется ротатабельностью, т. е. точки в матрице планирования подбираются так, что точность предсказания значений параметра оптимизации одинакова на равных расстояниях от центра эксперимента и не зависит от направления.

Для движения к точке оптимума нам нужна линейная модель $y = b_0 + b_1x_1 + b_2x_2$. Наша цель – найти по результатам эксперимента значения неизвестных

коэффициентов модели. До сих пор, говоря о линейной модели, мы не останавливались на важном вопросе о статистической оценке ее коэффициентов. Теперь необходимо сделать ряд замечаний по этому поводу. Можно утверждать, что эксперимент проводится для проверки гипотезы о том, что линейная модель $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ адекватна. Греческие буквы использованы для обозначения «истинных» генеральных значений соответствующих неизвестных. Эксперимент, содержащий конечное число опытов, позволяет только получить выборочные оценки для коэффициентов уравнения $y = b_0 + b_1 x_1 + \dots + b_k x_k$. Их точность и надежность зависят от свойств выборки и нуждаются в статистической проверке. Как производится такая проверка, будет показано ниже. А пока займемся вычислением оценок коэффициентов. Их можно вычислить по простой формуле

$$b_j = \frac{1}{N} \sum_{i=1}^N x_{ji} y_i, \quad j = 0, 1, \dots, k, \quad (2.11)$$

обоснование которой будет приведено ниже. Воспользуемся этой формулой для подсчета коэффициентов b_1 и b_2

$$b_1 = \frac{(-1)y_1 + (+1)y_2 + (-1)y_3 + (+1)y_4}{4}, \quad (2.12)$$

$$b_2 = \frac{(-1)y_1 + (-1)y_2 + (+1)y_3 + (+1)y_4}{4}. \quad (2.13)$$

Благодаря кодированию факторов расчет коэффициентов превратился в простую арифметическую процедуру. Для подсчета коэффициента b_1 используется вектор-столбец x_1 , а для b_2 – столбец x_2 . Остается неясным, как найти b_0 . Если уравнение $y = b_0 + b_1 x_1 + b_2 x_2$ справедливо, то оно верно и для средних арифметических значений переменных: $\bar{y} = b_0 + b_1 \bar{x}_1 + b_2 \bar{x}_2$. Но в силу свойства симметрии $\bar{x}_1 = \bar{x}_2 = 0$.

Следовательно, $\bar{y} = b_0$. Мы показали, что b_0 есть среднее арифметическое значений параметра оптимизации. Чтобы его получить, необходимо сложить все y и разделить на число опытов. Чтобы привести, эту процедуру в соответствие с формулой для вычисления коэффициентов, в матрицу планирования удобно ввести вектор-столбец фиктивной переменной x_0 , которая принимает во всех опытах значение +1. Это было уже учтено в записи формулы, где j принимало значения от 0 до k .

Теперь у нас есть все необходимое, чтобы найти неизвестные коэффициенты линейной модели

$$y = b_0 + b_1x_1 + b_2x_2 \quad (2.14)$$

Коэффициенты при независимых переменных указывают на силу влияния факторов. Чем больше численная величина коэффициента, тем большее влияние оказывает фактор. Если коэффициент имеет знак плюс, то с увеличением значения фактора параметр оптимизации увеличивается, а если минус, то уменьшается. Величина коэффициента соответствует вкладу данного фактора в величину параметра оптимизации при переходе фактора с нулевого уровня на верхний или нижний.

Иногда удобно оценивать вклад фактора при переходе от нижнего уровня к верхнему уровню. Вклад, определенный таким образом, называется вкладом фактора (иногда его называют основным или главным эффектом). Он численно равен удвоенному коэффициенту. Для качественных факторов, варьируемых на двух уровнях, основной уровень не имеет физического смысла. Поэтому понятие «эффект фактора» является здесь естественным.

Планируя эксперимент, на первом этапе мы стремимся получить линейную модель. Однако у нас нет гарантии, что в выбранных интервалах варьирования процесс описывается линейной моделью. Существуют способы проверки пригодности линейной модели (проверка адекватности).

Статистический вывод о пригодности (значимости) полученного в результате факторного анализа уравнения обычно проверяется в следующей последовательности:

Проводится общая проверка модели, целью которой является выяснение, объясняют ли переменные значимую долю изменения y . Определение значимости модели рекомендуется проводить по следующим методам (таблица 2.15).

Если регрессия не является значимой, то говорить больше не о чем.

В приведенном примере модель значима, т.к. вычисленный уровень значимости модели $p=0,000000 < 0,05$. Получаем уравнение вида

$$y(x_1, x_2, x_3, x_4) = 6,9 + 0,07 \cdot x_1 - 0,00035 \cdot x_2 - 2,08 \cdot x_3 + 0,00003 \cdot x_4. \quad (2.15)$$

Если регрессия оказывается значимой, то существует взаимосвязь между параметром y и переменными x_1, x_2, \dots, x_m . Однако остается неясно, каково влияние конкретных факторов x_1, x_2, \dots, x_m на исследуемую функцию y . Можно продолжить анализ, используя t -тесты для отдельных коэффициентов регрессии a_1, a_2, \dots, a_m с целью выяснить, насколько значимой является влияние той или иной переменной x на параметр y при условии, что все другие факторы x_k остаются неизменными. Проверку на адекватность коэффициентов уравнения рекомендуется проводить по следующим эквивалентным методам (таблица 2.16).

Т.к. вычисленные уровни значимости p -level для коэффициентов, стоящих при x_2 и x_4 меньше $0,05$, то они не значимы. К аналогичному выводу можно прийти, воспользовавшись t -критерием: $t_2(10) = -0,013 < 2,228$ и $t_3(10) = 1,44 < 2,228$. С учетом этого факта, пересчитаем уравнение множественной регрессии, выбрав в качестве зависимой (dependent) переменную y и независимые (independent) переменные x_1 и x_3 , коэффициенты при которых значимы.

Таблица 2.15 – Методы проверки модели, описываемой уравнением факторного анализа

Критерий Фишера	Использование уровня значимости α	Использование коэффициента детерминации R^2
<p>Проверяется нулевая гипотеза H_0 о равенстве полученных коэффициентов регрессии нулю: $a_1 = a_2 = a_m = 0$. Для этого рассчитанное системой значение F-критерия ($F_{расч}$), сравнивается с табличным значением $F_{табл}$, определяемым с использованием специальных таблиц по заданному уровню значимости (например, $\alpha=0,05$) и числу степеней свободы ($df1=m, df2=n-m-1$). Если выполняется неравенство $F_{расч} < F_{табл}$, то с уверенностью, например на 95 %, можно утверждать, что рассматриваемая зависимость является статистически значимой.</p>	<p>Если рассчитанное значение уровня значимости p больше, чем заданный уровень значимости α (например, $\alpha=0,05$), то полученный результат нужно трактовать как незначимый (для 95% вероятности). В том случае, когда величина $p < 0,05$, то вывод такой: это значимое уравнение с вероятностью 95%.</p>	<p>Рассчитанная системой величина $R^2_{расч}$ сравнивается с табличными (критическими) значениями $R^2_{крит}$, определяемым с использованием специальных таблиц по заданному уровню значимости (например, $\alpha=0,05$). Если окажется, что $R^2_{расч} > R^2_{крит}$, то с упомянутой степенью вероятности (95 %) можно утверждать, что анализируемая регрессия является значимой.</p>

Таблица 2.16 – Методы проверки адекватности коэффициентов уравнения

Использование t-критерия Стьюдента	Использование уровня значимости α
<p>Анализируемый коэффициент a_1, a_2, \dots, a_m считается значимым, если рассчитанное программой для него значение t-критерия по абсолютной величине превышает $t_{табл}$, определяемым с использованием таблиц по заданному уровню значимости (например, $\alpha=0,05$) и числу степеней свободы ($df = n - m - 1$).</p>	<p>Коэффициент регрессии a_1, a_2, \dots, a_m признается значимым, если рассчитанное программой для него значение уровня значимости p меньше (или равно) 0,05 (для 95%-ной доверительной вероятности).</p>

Список использованных источников

- 1 Сидняев, Н. И. Теория вероятностей и математическая статистика: учебник для бакалавров / Н. И. Сидняев. – Москва: Юрайт, ИД Юрайт, 2011. - 219 с.
- 2 Спирина, М. С. Теория вероятностей и математическая статистика: учебник для студ. учреждений сред. проф. образования / М. С. Спирина, П. А. Спирин. – Москва: ИЦ Академия, 2012. - 352 с.
- 3 Бычкова, С. Г. Социально-экономическая статистика: учебник для бакалавров / С. Г. Бычкова. – Москва: Юрайт, 2013. - 591 с.
- 4 Яковлева, А. В. Экономическая статистика: учебное пособие / А. В. Яковлева. – Москва: ИЦ РИОР, 2013. - 95 с.
- 5 Балдин, К. В. Общая теория статистики: учебное пособие / К. В. Балдин, А. В. Рукосуев. – Москва: Дашков и К, 2012. - 312 с.
- 6 Батракова, Л. Г. Теория статистики: учебное пособие / Л. Г. Батракова. – Москва: КноРус, 2013. - 528 с.
- 7 Громыко, Г. Л. Теория статистики: Практикум / Г. Л. Громыко. – Москва: НИЦ ИНФРА-М, 2013. - 238 с.
- 8 Ефимова, М. Р. Общая теория статистики: учебник / М. Р. Ефимова, Е. В. Петрова, В. Н. Румянцев. – Москва: ИНФРА-М, 2013. - 416 с.
- 9 Лысенко, С. Н. Общая теория статистики: учебное пособие / С. Н. Лысенко, И. А. Дмитриева. – Москва: ИД ФОРУМ, НИЦ ИНФРА-М, 2013. - 208 с.
- 10 Костин, В. Н. Теория эксперимента: учебное пособие / В. Н. Костин, В. В. Паничев; М-во образования и науки Рос. Федерации, Федер. гос. бюджет. образоват. учреждение высш. проф. образования «Оренбург. гос. ун-т», Каф. прогр. обеспечения вычисл. техники и автоматизир. систем. - Оренбург: Университет, 2014. - 212 с. : табл. - Библиогр.: с. 207-208. - Прил.: с. 209-212. - ISBN 978-5-4417-0415-1.
- 11 Килов, А. С. Планирование экспериментов и обработка экспериментальных данных: методические указания к практическому занятию для

студентов, обучающихся по программам высшего профессионального образования по направлению подготовки 150700.62 Машиностроение / А. С. Килов; М-во образования и науки Рос. Федерации, Федер. гос. бюджет. образоват. учреждение высш. проф. образования «Оренбург. гос. ун-т», Каф. материаловедения и технологии материалов. - Оренбург : ОГУ, 2014. - 35 с.

12 Годин, А. М. Статистика: учебник / А. М. Годин. - Москва: Дашков и К°, 2012. – 451 с.

13 Елисеева, И. И. Статистика : учебник для бакалавров / И. И. Елисеева. - Москва: Юрайт: ИД Юрайт, 2011. – 565 с.

14 Ниворожкина, Л. И. Статистика: учебник для бакалавров / Л. И. Ниворожкина. - Москва: Дашков и К°: Наука-Спектр, 2011. - 415 с.

15 Тумасян, А. А. Статистика промышленности: учебное пособие / А. А. Тумасян, Л. И. Василевская. - Минск: Новое знание. - Москва : Инфра-М, 2012. - 429 с.

16 Гмурман, В. Е. Теория вероятностей и математическая статистика: учебное пособие для бакалавров / В. Е. Гмурман. - Москва: Юрайт, 2013. - 479 с.

17 Горлач, Б. А. Теория вероятностей и математическая статистика: учебное пособие / Б. А. Горлач. - Санкт-Петербург: Лань, 2013. - 320 с.

18 Калинина, В. Н. Теория вероятностей и математическая статистика: учебник для бакалавров / В. Н. Калинина. - Москва: Юрайт, 2013. - 472 с.

19 Кобзарь, А. И. Прикладная математическая статистика. Для инженеров и научных работников / А. И. Кобзарь. - Москва: ФИЗМАТЛИТ, 2012. - 816 с.

20 Колемаев, В. А. Теория вероятностей и математическая статистика: учебник / В. А. Колемаев, В. Н. Калинина. - Москва: КноРус, 2013. - 376 с.

21 Кочетков, Е. С. Теория вероятностей и математическая статистика: учебник / Е. С. Кочетков, С. О. Смерчинская, В. В. Соколов. - Москва: Форум, НИЦ ИНФРА-М, 2013. - 240 с.

22 Краснов, М. Л. Вся высшая математика. Теория вероятностей. Математическая статистика. Теория игр: учебник / М. Л. Краснов, А. И. Киселев, Г. И. Макаренко. - Москва: ЛКИ, 2013. - 296 с.

23 Кремер, Н. Ш. Теория вероятностей и математическая статистика: учебник для студентов вузов / Н. Ш. Кремер. - Москва: ЮНИТИ-ДАНА, 2012. - 551 с.

24 Кричевец, А. Н. Математическая статистика для психологов: учебник для студ. учреждений высш. проф. образования / А. Н. Кричевец, А. А. Корнеев, Е. И. Рассказова. - Москва: ИЦ Академия, 2012. - 400 с.

25 Семенов, В. А. Теория вероятностей и математическая статистика: учебное пособие / В. А. Семенов. - Санкт-Петербург: Питер, 2013. - 192 с.