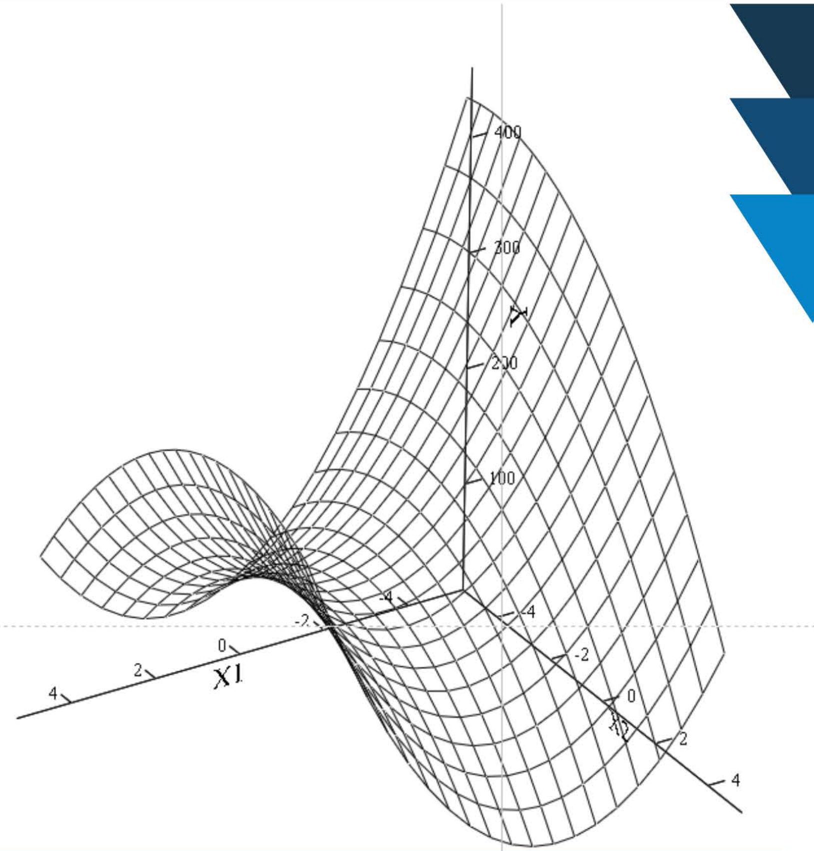
УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ ЛЕСОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

УГЛТУ

А. В. Мялицин



Методика планирования экспериментов и обработки их результатов при исследовании технологических процессов в деревообрабатывающей промышленности

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Уральский государственный лесотехнический университет»
(УГЛТУ)

А. В. Мялицин

Методика планирования экспериментов и обработки их результатов при исследовании технологических процессов в деревообрабатывающей промышленности

Учебно-методическое пособие

Екатеринбург
2023

УДК 674:001.891.5(075.8)
ББК 37.13в643я73
М99

Рецензенты:

кафедра лесного инжиниринга ФГБОУ ВО «Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева, д-р техн. наук, доцент *И. М. Еналеева-Бандура*;

В. В. Куртеев, директор ООО «Флекс»

М99 **Мялицин, Алексей Владимирович.**

Методика планирования экспериментов и обработки их результатов при исследовании технологических процессов в деревообрабатывающей промышленности : учебно-методическое пособие / А. В. Мялицин ; Министерство науки и высшего образования Российской Федерации ; Уральский государственный лесотехнический университет. – Екатеринбург : УГЛТУ, 2023. – 76 с.

ISBN 978-5-94984-877-7

В учебно-методическом пособии излагаются основы проведения теоретических и экспериментальных исследований, методика обработки полученных результатов, в том числе с применением офисного приложения *Microsoft Excel*. Приложения содержат основные статистические таблицы для определения критериев Стьюдента, Кохрена, Фишера, Пирсона.

Предназначено для обучающихся, осваивающих образовательные программы по направлению «Технология лесозаготовительных и деревоперерабатывающих производств».

Издается по решению редакционно-издательского совета Уральского государственного лесотехнического университета.

УДК 674:001.891.5(075.8)
ББК 37.13в643я73

ISBN 978-5-94984-877-7

ВВЕДЕНИЕ

В учебно-методическом пособии рассмотрены вопросы, касающиеся обработки результатов наблюдений, а также необходимые при планировании экспериментов.

При планировании эксперимента и последующей обработке результатов опытов методами регрессионного анализа требуется выполнение следующих предпосылок:

- результаты наблюдений должны представлять собой независимые нормально распределенные случайные величины;
- дисперсии этих случайных величин должны быть равны друг другу. Ниже подробно рассмотрены методы проверки этих утверждений.

Для экспериментатора-исследователя технологических процессов лесной промышленности излагаемые методы имеют и самостоятельную ценность. Во всех случаях, когда исследуемый параметр представляет собой случайную величину для получения характеристик этой величины (закона распределения, математического ожидания, дисперсии и т. д.) применимы излагаемые методы.

Сложность задач, решаемых при проведении научных исследований, обуславливает применение компьютерных технологий, поэтому для современного исследователя важно умение использовать различные пакеты прикладных программ, позволяющих проводить обработку экспериментальных данных и моделирования процессов.

1. СТАТИСТИЧЕСКИЕ ОЦЕНКИ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЯ

По результатам экспериментальных исследований технологических процессов изучается влияние целого ряда переменных величин, называемых *факторами*, на тот или иной процесс. К основным факторам будем относить все изучаемые факторы, а также те, которые служат для стабилизации процесса. Различные побочные и посторонние факторы желательно по возможности устранить. Некоторые из них можно с достаточной точностью учесть и тогда относим их к основным факторам. Основные факторы должны быть одинаковыми для всех наблюдений, посвященных исследованию какого-либо параметра, например шероховатости при резании древесины.

Экспериментальные исследования с постоянными, стабильными значениями основных факторов называются *однородными*. Однородность испытаний является одним из важнейших условий правильного применения статистических методов обработки наблюдений. Чтобы обеспечить однородность опытов, нужно каждую серию проводить на одной и той же установке, по неизменной методике, с одними и теми же измерительными приборами и исследователями, в реальный срок. При этом надо учесть, что многие факторы заметно меняются во времени и вызывают так называемый дрейф выходной измеряемой величины. Если избежать этого явления не удается, то его желательно учитывать как особый фактор. Все прочие неустранимые и неучитываемые факторы относят к случайным факторам.

Результат, который появился бы при воздействии одних только основных факторов, называется *идеальным результатом*. Вследствие воздействия случайных факторов реальный результат наблюдений всегда является случайной величиной. Отклонение реального результата от идеального называется *ошибкой наблюдений*. Она является случайной величиной и результатом действия случайных факторов.

Различают ошибки трех видов: систематические, грубые и случайные.

Систематическая ошибка повторяется по всей серии наблюдений. Она обычно связана с наличием неучтенных, но постоянных факторов. Устранение систематических ошибок помогает более тщательно учесть основные факторы, действующие в данном процессе.

Грубая ошибка связана с резким нарушением условий экспериментов или просчетом экспериментатора при отдельном наблюдении.

Грубые ошибки могут быть отброшены на основании проверки по специальным критериям, которые будут рассмотрены ниже.

Случайными ошибками называются те, которые появляются нерегулярно, причины возникновения которых неизвестны и которые невозможно учесть заранее.

В подавляющем большинстве реальных экспериментов наблюдения имеют распределение, близкое к нормальному, поэтому при обработке наблюдений первая основная гипотеза – это нормальность распределения результатов наблюдений.

Как правило, наблюдения производятся выборочным методом. Пусть имеется некоторая большая совокупность объектов или измерений, которые могут быть произведены над объектом. Эта совокупность называется *генеральной совокупностью*.

Она может быть конечной или бесконечной (например, при измерении непрерывной величины). Из этой совокупности извлекается выборка объемом n объектов или измерений. Выборка подвергается детальному исследованию, по результатам которого описывается вся генеральная совокупность, т. е. определяются статистические оценки, характер распределения и т. д. Ясно, что любое суждение о генеральной совокупности по выборке является случайным.

Главными статистическими оценками генеральной совокупности при условии нормальности распределения результатов наблюдений являются среднее арифметическое выборки $\overline{y}$ и дисперсия выборки σ^2. Среднее $\overline{y}$ является оценкой математического ожидания M_y генерального среднего), а дисперсия σ^2 – оценкой генеральной дисперсии D_y всей генеральной совокупности. Чем больше объем выборки, тем ближе выборочные $\overline{y}$ и σ^2 к M_y и D_y. Пусть имеется выборка $y\left(y_1, \ldots, y_i, \ldots, y_n\right)$, состоящая из n элементов (отдельных наблюдений). Среднее арифметическое и дисперсия выборки определяются по формулам

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{1}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2, \tag{2}$$

В формуле дисперсии (2) участвует величина $\bar{y}$, зависящая от элементов выборки, которая называется *связью*. Не нарушая этой одной связи, мы можем произвольно изменять лишь $(n-1)$ элементов выборки. Это число называется *числом степеней свободы выборки*:

$$f : f = n-1. \qquad (3)$$

Число степеней свободы входит в знаменатель формулы дисперсии.

Величина $\sigma = \sqrt{\sigma^2}$ называется *средним квадратичным отклонением выборки* или *выборочным стандартом*. Эта величина характеризует отклонения отдельных измерений от среднего и показывает размер вариации или колебания признака, обеспечиваемого при данной методике процесса и измерений его показателей.

Иногда в качестве оценки генерального среднего берется выборочная медиана. Для определения медианы нужно все элементы выборки расположить в возрастающем порядке. Средний элемент, т. е. такой, слева и справа от которого расположено одинаковое число элементов, и будет медианой. Если выборка имеет четное число элементов, то у нее два средних элемента, и в качестве медианы нужно брать их полусумму.

Важное значение в статистике имеют и другие показатели.

К ним относятся:

1) коэффициент вариации (в процентах):

$$\nu = \frac{\sigma}{\bar{y}} 100\,\%; \qquad (4)$$

2) средняя квадратичная ошибка среднего значения:

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}; \qquad (5)$$

3) показатель точности среднего значения в процентах:

$$\xi = \frac{\sigma_{\bar{y}}}{\bar{y}} 100\,\% = \frac{\nu}{\sqrt{n}}; \qquad (6)$$

4) ошибка среднего квадратичного отклонения:

$$\sigma_{\sigma} = \frac{\sigma}{\sqrt{2n}}. \qquad (7)$$

Коэффициент вариации и показатель точности в известной мере характеризуют надежность результатов наблюдений.

Все выборочные параметры являются случайными величинами [1]. Следовательно, их отклонения от генеральных параметров, т. е. погрешности и ошибки, также будут случайны. Мы можем лишь указать вероятность той или иной погрешности и тем самым выявить надежность (достоверность) каждого получаемого результата $+\varepsilon_0$.

Пусть случайная величина, определяющая ошибки результатов измерений, распределена по нормальному закону (рис. 1).

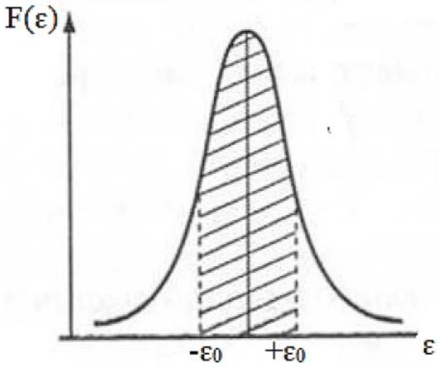


Рис. 1. Кривая нормального закона распределения

С некоторой вероятностью *p*, равной заштрихованной площади, ошибки лежат в границах $\pm\varepsilon_0$. Тогда истинный результат измерений v с вероятностью *p* будет находиться в пределах

$$\bar{y} - \varepsilon_0 \le \upsilon \le \bar{y} + \varepsilon_0. \qquad (8)$$

Вероятность (*p*) нахождения истинного результата, равного генеральному среднему, в указанных границах называется *уровнем достоверности* или *доверительной вероятностью*. Соответствующие доверительной вероятности пределы называются *доверительными границами*, а образуемый ими интервал – *доверительным интервалом*. С увеличением уровня достоверности до 100 % доверительный интервал растет до бесконечности, поэтому в зависимости от конкретных условий в качестве уровня достоверности (доверительной вероятности) берут 0,95; 0,99; 0,999. При известном среднем квадратичном отклонении (σ) доверительные интервалы строятся по следующей формуле:

$$\bar{y} - u\sigma \le \upsilon \le \bar{y} + u\sigma, \qquad (9)$$

где *u* – величина, для заданного уровня достоверности *p* выбираемая по табл. форме ниже.

Величина параметра *u*

p	*u*	*p*	*u*	*p*	*u*
0,999	3,89	0,98	2,33	0,88	1,28
0,999	3,29	0,96	2,05	0,70	1,04
0,998	3,09	0,95	1,96	0,60	0,84
0,995	2,81	0,90	1,64	0,50	0,67
0,990	2,58	0,85	1,44	0,20	0,25

Эта табл. форма справедлива для больших выборок объемом $n > 120$. Для малых выборок u можно определить по табл. У1 [2] с учетом числа степеней свободы $f = n - 1$. Здесь параметр u имеет t-критерий Стьюдента.

В табл. У1 [2] уровень достоверности p обозначен через α.

2. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Статистическая гипотеза – это некоторое предположение относительно свойств совокупности, из которой производилась выборка. К статистическим гипотезам можно отнести гипотезы о равенстве двух центров распределения (двух средних), о равенстве и однородности дисперсий, гипотезы о законе распределения, об адекватности уравнений регрессии и т. д.

Проверка статистической гипотезы – это процедура, по результатам которой гипотеза принимается или отбрасывается.

Проблема проверки статистических гипотез связана с такими распространенными задачами, как сравнительная оценка различных технологических процессов по их производительности, точности, экономичности или сравнение конструктивных особенностей машин и приборов.

В области теории эксперимента проверка статистических гипотез позволяет правильно оценить преимущества одной методики перед другой, выявить наиболее значимые факторы, влияющие на данное явление, а также убедить в пригодности (адекватности) полученного математического описания процесса.

Проверка гипотез часто сводится к сравнению статистических характеристик, по которым оцениваются параметры законов распределения. Строятся предположения, сводящиеся к некоторым утверждениям относительно значений параметров законов распределения. Из этих предположений выводятся следствия и рассматривается, насколько оправдываются они на практике. Следствия носят характер вероятностных суждений о поведении некоторых статистических характеристик при сделанных предположениях. Проверка заключается в вычислении этих характеристик по данным наблюдений и в сравнении их с тем, что было выведено на основе сделанных предположений. Такие характеристики называются *критериями проверки* или *критериями значимости*.

Принцип значимости заключается в том, что мы сосредотачиваем свое внимание в первую очередь на осуществившихся событиях, которые уже являются достоверными. Чем меньше расчетная вероятность уже осуществившегося события, тем больше его «неслучайность» и тем важнее эту «неслучайность» раскрыть.

Событие как бы сильнее приковывает к себе внимание наблюдателя, т. е. становится более значимым. Наибольшее значение вероятности, несовместимой со случайностью события, называют *уровнем значимости*. Иными словами, уровень значимости есть максимальная вероятность того, что случайное событие практически невозможно. Чем выше уровень значимости, тем он «жестче», т. к. при этом большее число событий нельзя рассматривать как случайные. Обычно принимаются уровни значимости q = 5 %, 2 %, 1 %, соответствующие события, которые считаются практически невозможными.

Построение критериев значимости связано со следующими распределениями:

- t-распределение Стьюдента;
- F-распределение Р. Фишера;
- χ^2-распределение К. Пирсона;
- λ-распределение А. Н. Колмогорова и т. д.

Согласно рис. 2 уровень значимости q отсекает в выбранном для для данного критерия распределения заштрихованную область, площадь которой равна $\frac{q}{100}$.

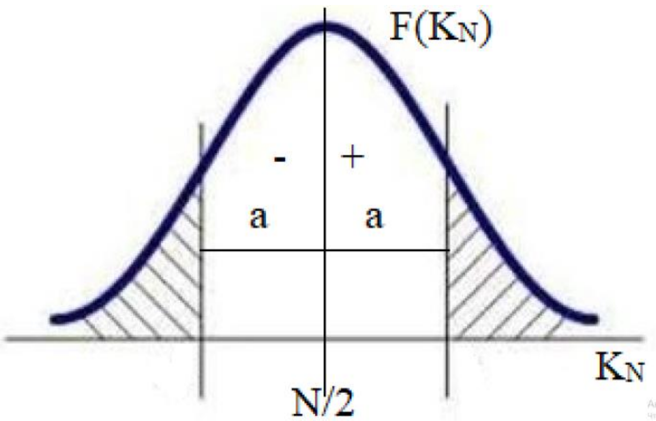


Рис. 2. Уровень значимости

Эта область называется *критической*. Она может быть односторонней и двухсторонней. Значения критерия, лежащие вне критической области, образуют область допустимых значений, площадь которой равна $1-\frac{q}{100}$. Таким образом, событие, соответствующее попаданию критерия значимости в область допустимых решений, признается практически достоверным, и гипотеза принимается. Если же значение критерия значимости, вычисленное по данным наблюдений, окажется в критической области, то мы отклоняем гипотезу, т. к. считаем, что при нашей гипотезе попадание критерия в критическую область практически невозможно, т. е. несовместимо с гипотезой.

Однако попадание критерия значимости в область допустимых значений не дает нам права категорически утверждать, что гипотеза полностью подтвердилась. Мы можем только заключить, что по данным нашей выборки (произведенных наблюдений) значение критерия не противоречит гипотезе. Дальнейшие наблюдения могут опровергнуть гипотезу, если критерий значимости хотя бы в одном из испытаний окажется в критической области. Мы видим, что суждение «гипотеза неверна» более категорично, чем суждение «гипотеза не отвергается, но и не утверждается, что она верна».

По этой причине, принимая решение о правильности гипотезы, мы можем допустить ошибку. *Ошибка первого рода* состоит в том, что отвергается гипотеза, которая на самом деле верна. Вероятность такой ошибки не выше уровня значимости q, т. е. мала. Она уменьшается с уменьшением уровня значимости. *Ошибка второго рода* состоит в том, что гипотеза принимается, а на самом деле она неверна. Уменьшение ошибки второго рода производится увеличением уровня значимости. Таким образом, уменьшение уровня значимости приводит к уменьшению ошибки первого рода и при этом к снижению чувствительности критерия значимости, а также к увеличению ошибки второго рода. Уровень значимости критерия проверки показывает вероятность ошибки первого рода, но не измеряет степень риска, связанного с ошибкой второго рода.

Для проверки гипотезы из всех возможных критериев стараются выбрать тот, у которого при заданном уровне значимости меньше вероятность ошибки второго рода. При этом необходимо, чтобы мы имели наибольшую вероятность попадания нашего критерия в критическую область, когда справедлива гипотеза, конкурирующая с проверяемой гипотезой. Эта вероятность носит название *мощности критерия*.

Рассмотрим более подробно некоторые задачи проверки статистических гипотез и уясним возможности применяемых критериев значимости.

Проверка гипотезы о значимости различия двух серий наблюдений

В экспериментальных исследованиях часто требуется сравнить два параллельных ряда наблюдений над объектами, принадлежащими к двум различным разновидностям, типам, сортам и т. д. При этом производится N опытов при планомерном изменении какого-либо фактора, различие в действии которого на каждую разновидность объектов хотят выяснить. В каждом опыте над парой объектов мы имеем два значения измеряемой величины: y_i и y'_i $(i = 1, \ldots, N)$. Вопрос заключается в том, можно ли считать различия между y_i и y'_i значимыми, существенными, т. е. связанными с различной способностью двух рассматриваемых объектов реагировать на изменения исследуемого фактора, или, наоборот, разности $y_i - y'_i = \eta_i$ следует отнести за счет случайного рассеяния значений измеряемой величины. Предположение об отсутствии существенного различия между параметрами сравниваемых генеральных совокупностей представляет собой *нулевую гипотезу*.

Мы допускаем, что случайные значения наблюдений величины в каждом опыте независимы друг от друга, последовательные N наблюдений независимы между собой и подчиняются одному и тому же закону распределения.

Отсюда следует, что разности $y_i - y'_i = \eta_i$ распределены симметрично около нуля, т. е. появления положительного и отрицательного отклонений равновероятны, т. е.

$$p("+\eta_i") = p("-\eta_i") = \frac{1}{2}. \tag{10}$$

При этом, естественно, не имеют абсолютные значения разностей, а разность $\eta_i = 0$ исключается из рассмотрения. Пусть число знаков «+» нашей последовательности составляет $K_{N(+)}$, тогда число «–» знаков составит $K_{N(-)} = N - K_{N(+)}$. Проверка нулевой гипотезы сведется к проверке значимости расхождения между частностью $W = \dfrac{K_{N(+)}}{N}$ появление знака «+» и вероятностью $p_{(+)} = \dfrac{1}{2}$.

Зададим уровень значимости $\frac{q}{100}$ и выберем критическую область больших по абсолютной величине отклонений. Это значит, что если $K_{N(+)}$ больше или меньше $\frac{N}{2}$ на некоторую величину «а», то мы попадаем в критическую область и отвергаем нулевую гипотезу. Здесь уровень значимости – это вероятность превышения абсолютного значения разности $\left|K_{N(+)} - \frac{N}{2}\right|$ некоторой величины «а» (рис. 2).

Таким образом,

$$\frac{q}{100} = p\left[\left|K_{N(+)} - \frac{N}{2}\right| \geq \alpha\right] \tag{11}$$

или

$$\frac{q}{100} = p\left[K_{N(+)} \geq \frac{N}{2} + \alpha\right] + \left[K_{N(+)} \leq \frac{N}{2} + \alpha\right]. \tag{12}$$

Если число m_N – меньшее из чисел $K_{N(+)}$ и $K_{N(-)}$, то вероятность того, что оно меньше половины опытов на величину «а», равна уровню значимости

$$p\left(m_N \leq \frac{N}{2} - \alpha\right) = \frac{q}{100}, \tag{13}$$

т. е. если m_N меньше критического числа $\overline{m}_N$ для заданного уровня значимости в процентах, то гипотеза об отсутствии расхождения между сериями наблюдений отвергается.

При $N > 90$ хорошее приближение критерия значимости m_N можно получить на основании нормального приближения и согласно биноминальному закону. Критическое число m_N можно вычислить из следующего выражения:

$$\overline{m}_N = \min\left(\frac{N-1}{2} \pm K\sqrt{N-1}\right), \tag{14}$$

где K = 1,2879 при q = 1 %; K = 0,9800 при q = 5 %; K = 0,8224 при q = 10 %.

Числа $\overline{m}_N$ приведены в прил. в табл. 1.

Представляет интерес число опытов N, необходимое для того, чтобы с надежностью $\beta \cdot 100\ \%$ мы получили значимый результат и забраковали неверную гипотезу (нулевую) $p_0 = \frac{1}{2}$, когда на самом деле верна гипотеза $p_0 \neq \frac{1}{2}$. Сведения о таких числах N для $\beta = 95\ \%$ приведены в табл. 1.

Таблица 1

Критическое число m_N в зависимости от уровня значимости

p	q, %					
	1		5		10	
	N	$\overline{m}_N$	N	$\overline{m}_N$	N	$\overline{m}_N$
0,45	1777	833	1297	612	1080	512
0,40	442	193	327	145	267	119
0,35	193	78	143	59	118	49
0,30	106	39	79	30	67	26
0,25	66	22	49	17	42	15
0,20	44	13	35	11	28	9
0,15	32	8	23	6	18	5
0,10	24	5	17	4	13	3
0,05	15	2	12	2	11	2

Из табл. видно: для того чтобы, например, в 95 % случаев отбросить гипотезу $p_0 = \frac{1}{2}$, когда на самом деле $p = 0{,}25$ с уровнем значимости $q = 5\ \%$, необходимо произвести 49 опытов, причем в этом случае $\overline{m}_{49} = 17$.

Из табл. 1 видно, что если значения p, конкурирующие с $p_0 = \frac{1}{2}$ близки к 1/2, то это различие можно уловить только при большом числе опытов N. Рассмотренный критерий получил название *критерия знаков*. Процедура его применения исключительно проста, измерения наблюдаемой величины могут производиться грубыми средствами, т. к. нас интересует лишь знак разности результатов наблюдений. Кроме того, он является непараметрическим, т. е. не зависит от формы распределения.

Сравнение дисперсий

Одной из важнейших задач статистической обработки наблюдений является сравнение двух или нескольких выборочных дисперсий. Основной выясняемый при этом вопрос – можно ли считать сравниваемые выборочные дисперсии оценками одной и той же генеральной дисперсии, т. е. равными, а расхождение между ними – случайным? Пусть имеется выборка в n наблюдений, дисперсия которой равна σ_1^2, а число степеней свободы $f_1 = n_1 - 1$. Имеется также вторая выборка в n_2 наблюдений, дисперсия которой равна σ_2^2, а число степеней свободы: $f_2 = n_2 - 1$. Выдвигается нулевая гипотеза о равенстве $\sigma_1^2 = \sigma_2^2$. Для того чтобы отвергнуть эту гипотезу, нужно доказать значимость расхождения между выборочными дисперсиями при выбранном уровне значимости q.

В качестве критерия значимости обычно используется критерий, основанный на F-распределении Р. Фишера.

F-распределением называется распределение случайной величины:

$$F = \frac{\sigma_1^2}{\sigma_2^2}. \tag{15}$$

Причем в качестве числителя берут большую из дисперсий. На рис. 3 показано распределение Фишера и двусторонняя критическая область, при попадании в которую гипотеза $\sigma_1^2 = \sigma_2^2$ отвергается. Здесь заштрихованные площади равны каждая $\frac{q}{2}$.

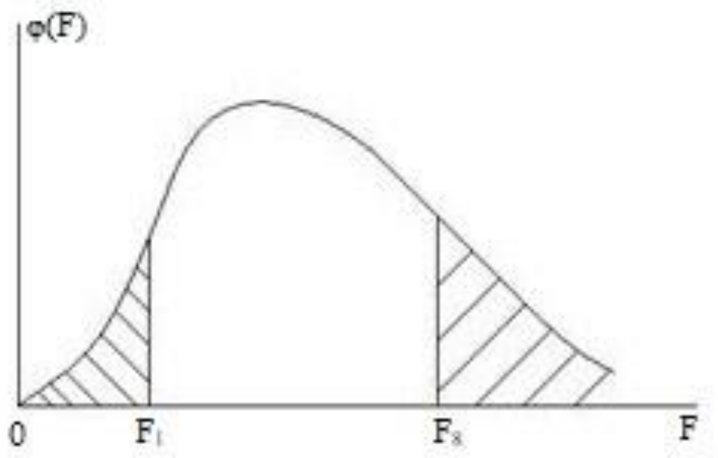


Рис. 3. Распределение Фишера

Пусть $F' = \dfrac{1}{F}$. Поскольку $F = \dfrac{\sigma_1^2}{\sigma_2^2}$ – имеет F_1-распределение со степенями свободы f_1 и f_2, то $F_1 = \dfrac{\sigma_2^2}{\sigma_1^2}$ имеет также F-распределение со степенями свободы f_2 и f_1. С помощью функции F' вероятность $\overline{p}$ того, что критерий значимости попадает в левую критическую область $P\left(F \le F_1\right)$, может быть выражена следующим образом:

$$\frac{p}{2} = P\left(F < F'\right) = p\left(\frac{1}{F} > \frac{1}{F_1}\right) = p\left(F' > \frac{1}{F}\right). \qquad (16)$$

Отсюда видно, что левая критическая точка F-распределения $\left(F < F_1\right)$ соответствует правой критической точке $\left(F' > \dfrac{1}{F}\right)$ или $\left(F' > F_1'\right)$ F-распределения.

Таким образом, при нахождении только правых точек для F и F' мы определяем правую и левую критические точки для F. В прил. в табл. 2 приведены только правые критические точки F-распределения. В этом случае гипотеза о равенстве дисперсий отбрасывается с уровнем значимости q, когда вычисленное значение F превосходит табличное для уровня значимости $\dfrac{q}{2}$.

Рассмотрим теперь вопрос о сравнении нескольких дисперсий $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$, имеющих число степеней свободы f_1, f_2, f_k. Требуется проверить однородность дисперсий, т. е. выяснить, являются ли числа σ_i^2 *(i = 1…K)* оценками одной и той же генеральной дисперсии. Сразу же напрашивается мысль использовать распределение Фишера, сравнивая дисперсии попарно. Такой путь ошибочен, т. к. незначимые различия, накапливаясь от пары к паре, могут быть вполне значимыми, хотя мы этого не заметим.

Такой ошибки можно избежать, сравнивая сразу самую большую и самую маленькую дисперсии. Если они отличаются незначимо, то и между промежуточными дисперсиями различий нет. Но и этот вывод лишь частично справедлив только при одинаковых объемах выборок.

Если все f_i одинаковы, т. е. объемы выборок равны, то более удобен критерий Кохрана:

$$G = \frac{\max \sigma_i^2}{\sum_{i=1}^{K} \sigma_i^2}\,. \tag{20}$$

В прил. в табл. 3 даны значения критерия G, отвечающие 1 % и 5 % уровням значимости.

3. ПРОВЕРКА ГИПОТЕЗ О ЗАКОНЕ РАСПРЕДЕЛЕНИЯ

Все статистические оценки опираются на нормальность наблюдаемого распределения и могут быть несправедливы в случае другого распределения. По этой причине применение этих оценок допустимо лишь при достаточной уверенности, что наблюдаемое распределение близко к нормальному. Гипотезу нормальности обычно проверяют непосредственно по наблюдениям (выборке), используя так называемые критерии согласия.

Основной принцип проверки по критериям согласия состоит в том, что заданная выборка сравнивается с некоторым заранее намеченным теоретическим распределением, т. е. мы задаем обычно нормальный характер теоретического распределения, а его среднюю $\overline{y}$ и дисперсию σ^2 подсчитываем из выборки, ибо других данных у нас нет.

Простейшие критерии согласия основаны на сравнении некоторых генеральных параметров предполагаемого теоретического распределения и выборочных параметров, полученных по результатам наблюдений. Более точными являются проверки по критериям Колмогорова и Пирсона.

Критерий А. Н. Колмогорова

Разобьем опытные данные выборки объемом n наблюдений на 8–20 одинаковых интервалов и построим выборочную эмпирическую функцию распределения. При этом необходимо для верхней границы y_i каждого интервала найти следующие величины:

а) эмпирическую частоту m, равную количеству наблюдений, попавших в данный интервал;

б) накопленную эмпирическую частоту m_N, равную сумме частот во всех предыдущих интервалах;

в) накопленную эмпирическую частотность $W_N = \dfrac{m_N}{n}$.

Затем построим график эмпирического распределения в интегральной форме $W_N = f\left(y_i\right)$.

Далее построим интегральную функцию предполагаемого теоретического распределения *F(y)*, вычислив значение по формуле:

$$F(y) = \frac{1}{2} + \Phi_0(z), \tag{21}$$

где $\Phi_0(z)$ – нормированная функция Лапласа (определяется по табл. 4 прил.),

$$z = \frac{y_i - \bar{y}}{\sigma}\ , \tag{22}$$

здесь $\bar{y}$ – средняя арифметическая выборка; σ – среднее квадратичное отклонение выборки (следует иметь в виду, что для отрицательных *z* функция Лапласа также отрицательна).

Теперь сравним отклонения эмпирических накопленных частостей W_{H_i} и теоретических вероятностей *F(y)* для верхней границы каждого интервала y_i, вычислив разности $W_{H_i} - F(y)$. Наибольшее отклонение $D = \max\left[W_{H_i} - F(y)\right]$ не превысит заданное число $\frac{\lambda}{\sqrt{n}}$ с вероятностью, имеющей λ-распределение. Определив $\lambda = D\sqrt{n}$ по табл. 5 прил. [3], находим соответствующее значение вероятности *p(λ)*, которое выполняет роль уровня значимости.

При пользовании критерия согласия А. Н. Колмогорова обычно берут высокие уровни значимости ($p \geq 0,2$). Если уровень значимости низкий (p < 0,05), то гипотеза нормальности отвергается. Графических построений можно избежать, если вычисление параметров эмпирического и теоретического распределений записывать по табл. форме ниже.

Расчет критерия Колмогорова

№ интервала	Верхняя граница интервала, y_i	Эмпирическая частота, m	Накопленная эмпирическая частота, m_N	Накопленная эмпирическая частость, W_N	z	Ф(z)	F(y)	$W_H - F(y)$
1	2	3	4	5	6	7	8	9

Критерий согласия χ^2 К. Пирсона

Критерий согласия К. Пирсона по сравнению с критерием А. Н. Колмогорова является более строгим и надежным для оценки степени различия двух сравниваемых рядов частот. При его помощи можно сравнивать эмпирический и теоретический или два эмпирических ряда. Этот критерий представляет собой сумму отношений квадратов разностей между частотами эмпирического и теоретического распределений:

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(m - nP_i\right)^2}{nP_i}, \tag{23}$$

где k – количество интервалов, на которые разбиты опытные данные; P_i – теоретические вероятности попадания опытных данных в i-ый интервал; nP_i – теоретические частоты попадания опытных данных в i-ый интервал (округляется до ближайших целых чисел); m – эмпирические частоты.

Вычисление теоретических вероятностей P_i производится по формуле:

$$P_i = \Phi_0\left(z_2\right) - \Phi_0\left(z_1\right), \tag{24}$$

где $z_1 = \dfrac{y_i^н - \overline{y}}{\sigma}$, $z_2 = \dfrac{y_i^в - \overline{y}}{\sigma}$, в которых $y_i^н$ – нижняя граница интервала, $y_i^в$ – верхняя граница интервала.

Результат вычислений для определения χ^2 можно свести в табличную форму ниже.

Расчет критерия Пирсона

№ интервала	Нижняя граница интервала, $y_i^н$	Верхняя граница интервала, , $y_i^в$	Эмпирические частоты, m	z_1	z_2	$\Phi_0\left(z_1\right)$	$\Phi_0\left(z_2\right)$	p_i	$p_i n$	$\left(m - p_i n\right)^2$	$\dfrac{\left(m - p_i n\right)^2}{p_i n}$
1	2	3	4	5	6	7	8	9	10	11	12

Суммируя показания последней графы табл. формы, получаем значение критерия χ^2. Далее выбираем уровень значимости q % для критерия. Пусть χ^2 обозначает q %-ный предел для закона χ^2 с (f – 3) степенями свободы. Этот предел можно определить по табл. 6 прил. Если гипотеза о нормальности выборки верна с уровнем значимости q %, то вычисленное значение χ^2 не попадает в область, т. е. $\chi^2 > \chi_q^2$, критерий попадает в критическую область, и гипотеза о нормальном законе распределения выборки отвергается.

Эмпирические и теоретические распределения для наглядности могут быть построены на графиках. При этом эмпирические частоты m и теоретические частоты nP_i, откладываемые по оси ординат, обычно относят к средним значениям интервала, которые откладываются по оси абсцисс.

Таким образом, проверка гипотезы о близости характера опытных данных к нормальному по критерию согласия χ^2 состоит из следующих этапов:

1) разбиение опытных данных на интервалы (не обязательно одинаковые);

2) определение эмпирических частот m попадания данных в i-ый интервал;

3) вычисление теоретических вероятностей P_i попадания данных в i-ый интервал;

4) вычисление теоретических частот nP_i попадания данных в i-ый интервал;

5) вычисление для каждого интервала величин $\frac{\left[m - P_i n\right]^2}{P_i n}$;

6) вычисление χ^2;

7) выбор уровня значимости q % и проверка условия $\chi^2 \le \chi_q^2$ по табл. (прил.);

8) построение графиков эмпирического и теоретического распределений.

В заключение следует ответить, что критерии А. Н. Колмогорова и К. Пирсона справедливы не только для нормального, но и для любого теоретического распределения, лишь бы оно было непрерывным.

4. ОТБРАСЫВАНИЕ ГРУБЫХ ИЗМЕРЕНИЙ

Одним из важнейших условий правильного применения статистических оценок является отсутствие грубых ошибок при наблюдениях. Все грубые ошибки должны быть найдены и исключены из рассмотрения в самом начале обработки наблюдений.

Самым надежным способом выявления грубых ошибок является тщательный анализ условий испытаний. При этом наблюдения, которые были проведены в нарушенных условиях опытов, должны отбрасываться. Однако на практике такой анализ можно провести не всегда. Обычно мы имеем дело с уже готовым цифровым материалом результатов наблюдений. Отдельные данные могут значительно отклоняться от остальных и вызывают у нас сомнение в их правильности. Отбрасывание этих данных должно производиться в результате строго научного анализа ряда наблюдений статистическим путем.

Как и всюду, мы будем исходить из нормального распределения ряда наблюдений. Каждая серьезная ошибка будет нарушать параметры нормального закона, т. е. будет нарушать однородность наблюдений, поэтому выявление глубоких измерений сводится к проверке гипотезы об однородности наблюдений.

Задача об однородности наблюдений формируется следующим образом: совместимы ли на уровне значимости q элементы выборки $y\left(y_1,\ \ldots,\ y_n\right)$ с гипотезой о том, что они извлечены из одной и той же генеральной совокупности, имеющей нормальное распределение. Если сомнение вызывает большое число элементов выборки, то их нужно выделить в отдельную выборку. Далее следует сравнить среднее арифметическое и дисперсии выделенной выборки и оставшейся и решить вопрос об однородности средних арифметических и дисперсий. Эта задача рассматривалась выше.

Чаще всего сомнительными бывают самый большой и самый малый элементы выборки так называемые крайние элементы.

Рассмотрим случай, когда генеральное среднее M_y и генеральная дисперсия D_y неизвестны. Пусть y_0 крайний элемент выборки. Величина

$$\tau = \frac{\left|y_0 - \overline{y}\right|}{\sigma} \qquad (25)$$

называется *максимальным относительным отклонением* и имеет специальное τ-распределение. С помощью этого распределения можно получить τ-критерий совместимости крайнего элемента с остальными. При этом не используется никаких сведений, кроме самой выборки.

Согласно τ-критерию крайнее значение y_0 отбрасывается с уровнем значимости q, если

$$\frac{\left|y_0 - \bar{y}\right|}{\sigma} > \tau_q. \qquad (26)$$

Значения τ_q для различных объемов выборки n приведены в табл. 2.

Таблица 2

Значение τ-критерия

n	Уровни значимости q, %		
	10	5	1
3	1,41	1,41	1,41
4	1,65	1,69	1,72
5	1,79	1,87	1,96
6	1,89	2,00	2,13
7	1,97	2,09	2,27
8	2,04	2,17	2,37
9	2,10	2,24	2,46
10	2,15	2,29	2,54
11	2,19	2,34	2,61
12	2,23	2,39	2,66
13	2,26	2,43	2,71
14	2,30	2,46	2,76
15	2,33	2,49	2,80
16	2,35	2,52	2,84
17	2,38	2,55	2,87
18	2,40	2,58	2,90
19	2,43	2,60	2,93
20	2,45	2,62	2,96
21	2,47	2,64	2,98
22	2,49	2,66	3,01
23	2,50	2,68	3,03
24	2,52	2,70	3,05
25	2,54	2,73	3,07

Таким образом, отбрасывание грубых измерений производится в следующем порядке:

1) вычисляется среднее $\bar{y}$ и дисперсия σ^2 с учетом всех элементов;

2) для крайних элементов, вызывающих сомнение, вычисляются максимальные относительные отклонения $\tau_{выч}$;

3) задается уровень значимости q;

4) крайние элементы выборки отбрасываются, если $\tau_{выч} > \tau_q$;

5) вычисляются среднее $\bar{y}$ и дисперсия выборки σ^2 без учета грубых измерений.

Иногда сомнение вызывают одновременно два или даже три элемента выборки. Тогда для каждого из сомнительных элементов вычисляется максимальное относительное отклонение τ. Исследование начинается с элемента, у которого τ наименьшее, а остальные сомнительные элементы временно отбрасываются. Вычисляем новые значения $\bar{y}$ и σ выборки без отброшенных элементов, а также новое значение τ' для оставшегося сомнительного элемента. Далее решаем вопрос об отбрасывании оставшегося сомнительного элемента с уровнем значимости q. Если $\tau' > \tau_q$, то оставшийся элемент выборки отбрасывается как грубое измерение. Тем более грубыми будут и остальные ранее отброшенные элементы. Если наименее сомнительный элемент не оказался ошибочным $\left(\tau' \le \tau_q\right)$, то его присоединяют к выборке и исследуют следующий по сомнительности элемент и т. д.

5. ОПРЕДЕЛЕНИЕ НЕОБХОДИМОГО ЧИСЛА НАБЛЮДЕНИЯ

Выборка из генеральной совокупности должна с достаточной точностью представлять параметры генеральной совокупности, т. е. быть репрезентативной. Представительность выборки зависит от ее объема.

Рассмотрим приближенный метод расчета необходимого числа наблюдений n для репрезентативной выборки, которую только предстоит произвести.

Здесь задача заключается в нахождении такого объема n выборки, чтобы вероятность p_0 отклонения выборочной средней $\bar{y}$ от генеральной средней M_y на величину, большую Δ, была очень мала. Тогда надежность получаемого результата (уровень достоверности) равна $p = 1 - p_0$. Чаще в технических исследованиях задача формулируется следующим образом: найти необходимое число измерений, чтобы с надежностью p % показатель точности среднего значения составлял величину не ниже, чем ξ %.

Метод исходит из условия, что дисперсия измеряемой величины постоянна и известна, а ее распределение является нормальным. Для определения дисперсии и проверки гипотезы нормальности ставится большая серия наблюдений (n > 120…150) для условий, в которых имеет место наибольшая вариация измеряемой величины. В результате этой серии определяются среднее квадратичное отклонение σ, коэффициент вариации υ.

Далее задаем надежность p, а также фиксируем допустимое отклонение Δ. Поскольку

$$\Delta = u\sigma_{\bar{y}}, \tag{27}$$

то с учетом (5) получим:

$$\Delta = u\sigma_{\bar{y}}, = u\frac{\sigma}{\sqrt{n}}, \tag{28}$$

здесь величина u зависит от надежности и определяется по табл. форме в гл. 1 (с. 8).

Из (28) получаем необходимое число наблюдений:

$$n = \frac{u^2\sigma^2}{\Delta^2} \tag{29}$$

или при заданном показателе точности:

$$n = \frac{\nu^2}{\xi^2}\ . \tag{30}$$

Следует отметить, что формула (30) менее предпочтительна, т. к. требует постоянства коэффициента вариации ν и не позволяет оценить надежность получаемого результата.

Формулу (29) следует применять для больших генеральных совокупностей и для уровней достоверностей, равных или меньших $\overline{\overline{p}} \leq 0{,}9$. При больших уровнях достоверности величина зависит от n.

Для подобной ситуации нами предлагается метод, в основу которого положено итеративное приближение к необходимому объему выборки.

Метод включает в себя следующие этапы:

1) вычислить n по формуле (29) с использованием табл. формы в гл. 1 (с. 8) для определения u;

2) вычислить число степеней свободы выборки $f = n - 1$ и определить новое u по табл. У1 [2] для заданного уровня достоверности $\overline{\overline{p}} = \alpha$;

3) вычислить новое n, новое число степеней свободы f и вновь определить u по табл. У1 [2];

4) опять вычислить n и т. д.

Вычисления повторять до тех пор, пока последующее n будет отличаться от предыдущего не более, чем на единицу.

В случае меньших объемов генеральной совокупности, равных N, необходимое число измерений n определяется по формуле

$$n = \frac{u^2\sigma^2 N}{(N-1)\Delta^2 + u^2\sigma^2}. \tag{31}$$

Здесь также может быть использован метод итерации для уточнения n при больших надежностях p.

Рекомендуемые формулы (29) и (31) пригодны лишь для определения объема выборки, т. е. результатов опытов.

Для проверки статистических гипотез расчет объема выборки весьма сложен и здесь не рассматривается. Ориентировочно следует принимать объем выборки для проверки статистических гипотез не менее 120…200 наблюдений.

6. ПРИМЕР СТАТИСТИЧЕСКОЙ ОБРАБОТКИ РЕЗУЛЬТАТОВ ОПЫТА ПРИ ИССЛЕДОВАНИИ ШЕРОХОВАТОСТИ ПОВЕРХНОСТИ ПИЛОМАТЕРИАЛОВ

Применим изложенные методы для обработки результатов эксперимента по исследованию шероховатости поверхности пиломатериалов.

В первоначальном опыте выполнено 140 замеров неровностей разрушения. На основании этих данных (см. табл. 3, опыт № 25) определяем необходимое число наблюдений в каждом опыте с целью получения репрезентативной выборки согласно гл. 5. Для этого вычислим для опыта № 25 среднее значение $\bar{y}$ по формуле (1) и оценку дисперсии σ^2 по формуле (2):

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\left(y_1 + y_2 + \ldots y_n\right) = \frac{1}{140}\left(760 + 760 + 460 + \ldots + 452\right) = 749,7,$$

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \frac{1}{n-1}\left[\left(y_1 - \bar{y}\right)^2 + \left(y2 - \bar{y}\right)^2 + \ldots\left(y_n - \bar{y}\right)^2\right] =$$

$$= \frac{1}{140-1}\left[\left(760 - 749,7\right)^2 + \left(760 - 749,7\right)^2 + \ldots\left(452 - 749,7\right)^2\right] = 40916,9.$$

Найдем оценку среднеквадратического отклонения:

$$\sigma = \sqrt{\sigma^2} = \sqrt{40916,9} = 202,3.$$

Задаем допустимое отклонение $\Delta = 100$ и величину $u = 2,05$, которая соответствует надежности $p = 0,96$. Тогда по формуле (29) получаем необходимое число наблюдений n:

$$n = \frac{u^2\sigma^2}{\Delta^2} = \frac{2,05^2 40916,9}{100^2} = 17,2 = 18.$$

После этого расчета были проделаны остальные 24 опыта (табл. 3, опыт 1 + 24). В каждом из этих этапов было проведено 20 замеров.

Проведем теперь для каждого опыта вычисления величин $\bar{y}, \sigma^2, \sigma$, точно так же, как это было сделано для опыта № 25. Результаты вычислений приведены в табл. 4.

Таблица 3

Результаты опыта

№ опыта	Результаты наблюдений						Число замеров	Среднее арифметическое
1	730	730	730	730	730	730	20	541,0
	460	460	460	460	460	460		
	460	460	460	460	460	460		
	460	460						
2	1100	1497	1230	829	1000	880	20	968,9
	646	878	573	829	1132	1 013		
	852	1220	1150	890	695	988		
	1050	926						
3	1115	1278	1090	1060	1078	1152	20	1086,4
	1152	1370	1400	1384	1039	1017		
	1247	908	770	831	807	770		
	924	1335						
4	1213	1386	595	595	595	1 565	20	1017,4
	983	1597	975	732	1250	1 160		
	1050	1160	1050	594	693	742		
		1475	937					
5	462	623	607	591	558	671	20	561,0
	624	462	446	543	462	429		
	470	712	712	470	550	567		
	631	630						
6	522	650	434	670	488	495	20	605,0
	568	582	695	636	569	569		
	662	682	782	755	695	609		
	521	515						
7	571	585	715	417	630	548	21	583,0
	651	566	599	597	463	597		
	503	558	608	707	603	690		
	716	440	480					
8	800	990	890	850	1750	800	20	1066,0
	1000	1000	1450	1100	730	1 310		
	1010	1010	1080	1260	680	1 080		
	1440	1090						
9	1405	1047	1572	1004	897	1 047	20	1029,4
	1047	1106	834	977	980	980		
	982	970	975	975	975	970		
	970	875						
10	432	605	461	794	438	552	20	638,5
	998	500	560	570	542	487		
	665	893	547	937	830	492		
	855	612						

Продолжение табл. 3

№ опыта	Результаты наблюдений						Число замеров	Среднее арифметическое
11	670	670	670	670	670		20	575,5
	544	544	544	544	544			
	544	544	544	544	544			
	544	544	544	544	544			
12	627	506	579	510	619		20	566,0
	619	619	619	540	540			
	540	437	437	857	540			
	540	540	540	540	570			
13	820	810	1080	1070	1380		20	1063,5
	1400	1230	1120	930	1040			
	980	1420	1270	680	670			
	1020	1050	1170	1080	1050			
14	1213	1386	595	595	595		20	1009,6
	1565	983	1557	975	732			
	1250	1160	1050	594	693			
	742	855	1475	937	1240			
15	445	503	524	701	460		20	670,1
	635	825	512	657	652			
	554	584	730	752	854			
	766	810	817	803	817			
16	950	980	1000	860	820		20	1062,5
	1070	760	1920	890	1060			
	700	650	750	820	1620			
	980	1150	1050	1300	1920			
17	460	460	460	460	460		20	425,0
	460	460	460	460	460			
	390	390	390	390	390			
	390	390	390	390	390			
18	1184	1080	1550	1402	1410		20	1358,9
	1830	1290	1177	1125	1430			
	1115	1410	1020	1445	1125			
	1168	2170	1240	2030	976			
19	578	726	695	741	718		20	718,5
	523	884	967	906	859			
	820	546	492	732	710			
	592	718	640	492	1030			
20	995	835	616	797	1105		20	900,4
	679	925	1000	1195	1255			
	728	752	640	940	1442			
	616	880	820	962	825			

Окончание табл. 3

№ опыта	Результаты наблюдений						Число замеров	Среднее арифметическое
21	670	747	824	584	550		20	695,2
	524	696	679	842	567			
	575	626	558	670	901			
	987	944	250	1135	575			
22	690	790	790	850	968		20	817,0
	1040	731	858	723	562			
	867	952	748	824	645			
	748	771	765	918	1100			
23	864	882	603	631	835		20	755,2
	660	548	902	586	697			
	826	864	929	992	780			
	632	771	678	642	781			
24	750	710	700	760	735		20	730,0
	725	710	750	720	740			
	765	695	715	745	712			
	748	730	740	720	730			
25	760	940	1110	766	615	502	140	749,7
	760	1010	850	790	535	517		
	460	454	847	795	500	720		
	460	685	844	892	740	730		
	430	910	942	1120	758	619		
	651	950	1236	835	785	623		
	668	910	545	830	802	510		
	602	554	852	900	1034	702		
	605	558	860	910	1146	800		
	545	428	902	1210	801	767		
	521	660	910	626	845	862		
	750	675	910	560	842	1074		
	708	618	560	552	960	1110		
	629	618	428	560	1260	825		
	605	535	485	860	613	825		
	725	510	652	864	1070	860		
	775	734	675	560	412	952		
	780	741	602	565	1160	1264		
	880	717	610	495	452	660		
	1117	740	523	456	552	950		
	1180	710	503	680	684	960		
	816	752	730	692	570	960		
	844	890	710	630	560	960		
	1060	452						

Таблица 4

Основные показатели выборки

№ опыта	Кол-во опытов	Среднее арифметическое, $\bar{y}$	Дисперсия, σ^2	Средн. квадратич. ошибка, σ	Средн. ошибка средн. арифм, $\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$	Коэффициент вариации, $\nu = \frac{\sigma}{\bar{y}} 100\ \%$	Показатель точности $\xi = \frac{\sigma_{\bar{y}}}{\bar{y}} 100\ \%$
1	20	541,0	16114,7	126,9	28,39	23,5	5,2
2	20	968,9	47874,6	218,8	48,93	22,6	5,0
3	20	1086,4	42664,5	206,6	46,19	19,0	4,3
4	20	1017,4	110670,2	332,7	74,39	32,7	7,3
5	20	561,0	8173,5	90,4	20,22	16,1	3,6
6	20	605,0	8842,1	94,0	21,03	15,5	3,5
7	20	588,2	7420,8	86,1	19,26	14,6	3,3
8	20	1066,0	71383,2	267,2	59,74	25,1	5,6
9	20	1029,4	29043,1	170,4	38,11	16,6	3,7
10	20	638,5	31964,8	178,8	39,98	28,0	6,3
11	20	575,5	3133,4	56,0	12,52	9,7	2,2
12	20	566,0	7644,7	87,4	19,55	15,4	3,5
13	20	1063,5	46266,1	215,1	48,10	20,2	4,5
14	20	1009,6	111119,8	333,3	74,54	33,0	7,4
15	20	670,1	18439,6	135,8	30,36	20,3	4,5
16	20	1062,5	134472,4	366,7	82,00	34,5	7,7
17	20	425,0	1289,5	35,9	8,03	8,4	1,9
18	20	1358,9	105969,6	325,5	72,79	24,0	5,4
19	20	718,5	24485,7	156,5	34,99	21,8	4,9
20	20	900,4	48970,0	221,3	49,48	24,6	5,5
21	20	695,2	39741,4	199,4	44,58	28,7	6,4
22	20	817,0	17322,8	131,6	29,43	16,1	3,6
23	20	755,2	16686,8	129,2	28,88	17,1	3,8
24	20	730,0	389,4	19,7	4,41	2,7	0,6
25	140	749,7	40916,9	202,3	17,10	27,0	2,3

Рассмотрим задачу обнаружения промахов при помощи τ-критерия. Для нашего объема выборки $n = 20$ и для 5 %-ного уровня значимости $q = 0{,}05$ по табл. 2 находим значение $q_\tau = 2{,}62$. Из анализа табл. 3 под сомнение взяты самый большой элемент выборки в опытах 8 (1750), 16 (1920) и 18 (2170) и самый малый элемент в опыте 16 (650).

Для этих элементов вычисляем по формуле (25) максимальные относительные отклонения τ. Величины $\bar{y}, \sigma$ берутся из табл. 4 (например, для опыта 8 $\bar{y} = 1066, \sigma = 267,2$). Тогда

$$\tau = \frac{\left|y_0 - \bar{y}\right|}{\sigma} = \frac{\left|1750 - 1066\right|}{267,2} = 2,56.$$

Аналогичные вычисления проведены для остальных сомнительных элементов. Результаты вычислений:

№ опыта	τ
8	2,56
16 (max)	2,34
16 (min)	1,12
18	2,49.

Согласно правилу крайние элементы выборки отбрасываются, если $\tau > \tau_q$.

Из полученных нами данных видно, что для всех подозрительных случаев $\tau < \tau_q$, следовательно, данные элементы выборки не являются промахами.

Проверим гипотезу об однородности ряда дисперсий с помощью критерия Кохрана (критерий применим, т. к. в 1–24 опытах объемы выборок равны). Воспользуемся формулой (20). В нашем случае σ^2 максимальная (в 16 опыте): $\sigma_{16}^2 = 134472,4$. Тогда

$$G = \frac{\max \sigma_i^2}{\sum_{i=1}^{K} \sigma_i^2} = \frac{134472,4}{950082,7} = 0,142.$$

Зададим 0,5 %-ным уровень значимости. По величине $f = n - 1 = 19$ (n – объем выборки) и величине K = 24 (K – число выборок) из табл. 3 прил. получим: $G_{табл} = 0,154$.

Полученный результат $G < G_{табл}$ позволяет принять гипотезу об однородности дисперсий с 0,5 %-ным уровнем значимости.

Проверка гипотез о законе распределения

Гипотезу нормальности данного распределения проведем с помощью критериев Колмогорова и Пирсона.

Применение критерия Колмогорова проиллюстрируем на опыте № 25 с числом наблюдений n = 140. С этой целью заполним табл. 5.

Таблица 5

Результаты вычисления критерия Колмогорова

№ интервала	Верхняя граница интервала	Число элементов в интервале, m	$m_н$	$W_н$	z	$\Phi_0(z)$	$F(y)$	$W_н - F(y)$
1	450	4	4	0,029	–1,48	–0,4306	0,0694	–0,041
2	500	9	13	0,093	–1,23	–0,3907	0,1093	–0,016
3	550	11	24	0,171	–0,99	–0,3389	0,1611	0,010
4	600	11	35	0,250	–0,74	–0,2704	0,2296	0,020
5	650	14	49	0,350	–0,49	–0,1879	0,3121	0,038
6	700	11	60	0,429	–0,25	–0,0987	0,4013	0,027
7	750	14	74	0,529	0,00	0,0000	0,5	0,029
8	800	12	86	0,614	0,25	0,0987	0,5987	0,016
9	850	13	99	0,707	0,50	0,1915	0,6915	0,016
10	900	10	109	0,779	0,74	0,2704	0,7704	0,008
11	950	10	119	0,850	0,99	0,3389	0,8389	0,011
12	1000	5	124	0,886	1,24	0,3925	0,8925	–0,007
13	1050	2	126	0,900	1,48	0,4306	0,9306	–0,031
14	1100	3	129	0,921	1,73	0,4582	0,9582	–0,037
15	1150	5	134	0,957	1,98	0,4761	0,9761	–0,019
16	1200	2	136	0,971	2,23	0,4871	0,9871	–0,016
17	1250	2	138	0,986	2,47	0,4932	0,9932	–0,007
18	1300	2	140	1,000	2,72	0,4967	0,9967	0,003
–	n = 140		–	–	–	$\max\left[W_н - F(y)\right]$		0,041
–	–	–	–	–	–	λ		0,483

Опытные данные нашей выборки разобьем на 18 интервалов. Верхние границы каждого интервала приведены во второй графе табл. 5. В следующей графе вписана эмпирическая частота m, равная количеству наблюдений, попавших в данный интервал (в частности, для 1-го интервала находим все наблюдения меньше 450, см. табл. 3).

Их четыре: 430, 412, 428, 428. Следовательно, $m_1 = 4$. Для 2-го интервала находим все наблюдения меньше 500 и больше 450. Их девять: $m_2 = 9$ и т. д.

Далее вписываем накопленную эмпирическую частоту $m_н$, равную сумме эмпирических частот во всех предыдущих интервалах (например, для 4-го интервала $m_{н4} = m_1+m_2+m_3+m_4$ = 4+9+11+11=35

Затем вписываем наполненную эмпирическую частотность $W_н$, равную величине $m_н$, деленной на число наблюдений *n*. Для 4-го интервала:

$$W_н = \frac{m_{н4}}{n} = \frac{35}{140} = 0,25.$$

В 6-ю графу вписываем величину *z*, вычисляемую по формуле (22). В ней $\overline{y}, \sigma$ берутся из табл. 3 для 25-го опыта: $\overline{y} = 749,7; \sigma = 202,3$. Для 4-го интервала:

$$z = \frac{600 - 749,7}{202,3} = -0,74.$$

В 7-ю графу табл. 5 вписываем величину $\Phi_0(z)$, которую находим из табл. 4 прил. по значениям *z*.

Далее вписываем величину $F(y)$, которую вычисляем по формуле (21).

Например, для 4-го интервала:

$$F(y) = \frac{1}{2} + \Phi_0(-0,74) = \frac{1}{2} - 0,2704 = 0,2296.$$

Наконец, последнюю графу табл. 5 занимает величина $W_н - F(y)$. Для 4-го интервала: $W_н - F(y) = 0,25 - 0,2296 = 0,02$.

Из величин, записанных в последней графе табл., выберем максимальную. Очевидно, что эта величина *D* = 0,041. Найдем значение $\lambda = D\sqrt{n} = 0,041 \cdot \sqrt{140} = 0,483$. Для полученного λ из табл. 5 прил. найдем значение *p*(λ) = 0,965. Мы получили *p*(λ) > 0,2. Это значит, что у нас нет оснований отвергнуть гипотезу о нормальности распределения.

Более предпочтительным является критерий Пирсона. Для его применения сначала заполним табл. 6. Как и для критерия Колмогорова опытные данные выборки разобьем на 18 интервалов и в 1–3-ей графах запишем нижние границы интервалов и эмпирические частоты (табл. 6). Поскольку разбиения на интервалы для критерия Колмогорова и Пирсона одинаковы, то 3-я и 4-я графы (табл. 6) совпадают с графами 2-ой и 3-й (табл. 5) соответственно.

Таблица 6

Результаты расчетов критерия Пирсона

№ интервала	Нижняя граница интервала	Верхняя граница интервала	Число элементов в интервале, *m*	z_1	z_2	$\Phi_0(z_1)$	$\Phi_0(z_2)$	p_i	$p_i n$	$(m - p_i n)^2$	$\frac{(m - p_i n)^2}{p_i n}$
1	400	450	4	–1,73	–1,48	–0,4582	–0,4306	0,0276	3,86	0,0185	0,005
2	450	500	9	–1,48	–1,23	–0,4306	–0,3907	0,0399	5,59	11,6554	2,087
3	500	550	11	–1,23	–0,99	–0,3907	–0,3389	0,0518	7,25	14,0475	1,937
4	550	600	11	–0,99	–0,74	–0,3389	–0,2704	0,0685	9,59	1,9881	0,207
5	600	650	14	–0,74	–0,49	–0,2704	–0,1879	0,0825	11,55	6,0025	0,520
6	650	700	11	–0,49	–0,25	–0,1879	–0,0987	0,0892	12,49	2,2141	0,177
7	700	750	14	–0,25	0,00	–0,0987	0	0,0987	13,82	0,0331	0,002
8	750	800	12	0,00	0,25	0	0,0987	0,0987	13,82	3,3051	0,239
9	800	850	13	0,25	0,50	0,0987	0,1915	0,0928	12,99	0,0001	0,000
10	850	900	10	0,50	0,74	0,1915	0,2704	0,0789	11,05	1,0941	0,099
11	900	950	10	0,74	0,99	0,2704	0,3389	0,0685	9,59	0,1681	0,018
12	950	1000	5	0,99	1,24	0,3389	0,3925	0,0536	7,50	6,2700	0,836
13	1000	1050	2	1,24	1,48	0,3925	0,4306	0,0381	5,33	11,1156	2,084
14	1050	1100	3	1,48	1,73	0,4306	0,4582	0,0276	3,86	0,7465	0,193
15	1100	1150	5	1,73	1,98	0,4582	0,4761	0,0179	2,51	6,2200	2,482
16	1150	1200	2	1,98	2,23	0,4761	0,4871	0,011	1,54	0,2116	0,137
17	1200	1250	2	2,23	2,47	0,4871	0,4932	0,0061	0,85	1,3133	1,538
18	1250	1300	2	2,47	2,72	0,4932	0,4967	0,0035	0,49	2,2801	4,653
–	–	–	–	–	–	–	–	–	–	χ^2	17,214

Для каждого интервала найдем величины $z_1 = \dfrac{y_i^н - \overline{y}}{\sigma}$, $z_1 = \dfrac{y_i^в - \overline{y}}{\sigma}$, где $y_i^н$ – нижняя граница интервала, $y_i^в$ – верхняя граница интервала, $\overline{y}$ – оценка математического ожидания, а σ – оценка среднеквадратического отклонения нашей выборки. Из табл. 4 имеем: $\overline{y} = 749,7; \sigma = 202,3.$ Например, для второго интервала $z_1 = \dfrac{450 - 749,7}{202,3} = -1,48$; $z_2 = \dfrac{500 - 749,7}{202,3} = -1,23.$ Значения z_1 и z_2 заносим в 5-ю и 6-ю графы табл. 6. По табл. 4 прил. для каждого интервала находим $\Phi_0(z_2)$, $\Phi_0(z_1)$ и по формуле (24) вычисляем P_i. Так, для второго интервала $\Phi_0(z_1) = \Phi_0|-1,48| = -0,4306$, $\Phi_0(z_2) = \Phi_0|-1,23| = -0,3907.$ Значения $\Phi_0(z_2)$, $\Phi_0(z_1)$ находятся в 7-ой и 8-ой графах табл. 6 $P_i = \Phi_0(z_2) - \Phi_0(z_2) = = -0,3907 + 0,4306 = 0,0399.$ Найденные p_i заносим в 9-ю графу таблицы 6. В 10-ю графу записываем значения $p_i n$(n = 140 – число наблюдений). В 11-й графе приведены результаты вычислений величины $(m - p_i n)^2$ (для 2-го интервала эта величина равна $(9 - 5,59)^2 = 11,6554$). В 12-й графе приведены значения $\dfrac{(m - p_i n)^2}{p_i n}$. Для 2-го интервала имеем: $\dfrac{11,6554}{5,59} = 2,087$.

Суммируя показания последней графы, получаем значение

$$\chi^2 = \sum_{i=1}^{18} \frac{(m - p_i n)^2}{p_i n} = 0,005 + 2,087 + \ldots + 4,653 = 17,214.$$

Зададим уровень значимости q = 5 %. Зная число интервалов l, на которое разбиты опытные данные (у нас l = 18), вычислим число степеней свободы f = l – 3 = 15. По величинам q и f из табл. 6 прил. находим значение χ_q^2. В нашем случае $\chi_q^2 = 25$. Как видно, $\chi^2 < \chi_q^2$, поэтому гипотеза о нормальности данного распределения может быть принята.

7. ПРИМЕНЕНИЕ EXCEL ПРИ ПЕРВИЧНОЙ ОБРАБОТКЕ РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА

В данной главе показано применение программы *Excel* для анализа результатов эксперимента.

Построение гистограммы и кривой распределения

Запишем исходные данные опыта № 25 (рис. 4).

	A	B	C	D	E	F
1	760	940	1110	766	615	502
2	760	1010	850	790	535	517
3	460	454	847	795	500	720
4	460	685	844	892	740	730
5	430	910	942	1120	758	619
6	651	950	1236	835	785	623
7	668	910	545	830	802	510
8	602	554	852	900	1034	702
9	605	558	860	910	1146	800
10	545	428	902	1210	801	767
11	521	660	910	626	845	862
12	750	675	910	560	842	1074
13	708	618	560	552	960	1110
14	629	618	428	560	1260	825
15	605	535	485	860	613	825
16	725	510	652	864	1070	860
17	775	734	675	560	412	952
18	780	741	602	565	1160	1264
19	880	717	610	495	452	660
20	1117	740	523	456	552	950
21	1180	710	503	680	684	960
22	816	752	730	692	570	960
23	844	890	710	630	560	960
24	1060	452				

Рис. 4. Исходные данные для анализа

Определим объем выборки. Для этого используем функцию «СЧЕТ».

В ячейку J26 вводим «”=СЧЕТ(A1:F24)”» (рис. 5).

	A	B	C	D	E	F	I	J
1	760	940	1110	766	615	502		
2	760	1010	850	790	535	517		
3	460	454	847	795	500	720		
4	460	685	844	892	740	730		
5	430	910	942	1120	758	619		
6	651	950	1236	835	785	623		
7	668	910	545	830	802	510		
8	602	554	852	900	1034	702		
9	605	558	860	910	1146	800		
10	545	428	902	1210	801	767		
11	521	660	910	626	845	862		
12	750	675	910	560	842	1074		
13	708	618	560	552	960	1110		
14	629	618	428	560	1260	825		
15	605	535	485	860	613	825		
16	725	510	652	864	1070	860		
17	775	734	675	560	412	952		
18	780	741	602	565	1160	1264		
19	880	717	610	495	452	660		
20	1117	740	523	456	552	950		
21	1180	710	503	680	684	960		
22	816	752	730	692	570	960		
23	844	890	710	630	560	960		
24	1060	452						
25								
26							Объем выборки	=СЧЁТ(A1:F24)
27								

Рис. 5. Определение объема выборки

Определим минимальное и максимальное значения выборки. Для этого используем встроенные функции МИН и МАКС. Для этого в ячейку J27 вводим формулу «”=МИН(A1:F24)”» и «”=МАКС(A1:F24)”». Результаты вычислений приведены на рис. 6.

J28 =МАКС(A1:F24)

	I	J
25		
26	Объем выборки	140
27	минимальное значение	412
28	максимальное значение	1264

Рис. 6. Вычисление максимального и минимального значений выборки

Необходимое число интервалов k определяется по формуле

$$k = 1 + 3,2\lg n, \qquad (32)$$

где n – объем выборки.

В ячейку J29 запишем выражение «”=1+3,2*LOG10(J26)”». Полученное значение необходимо округлить в большую сторону до целого. Для этого используется функция ОКРУГЛВВЕРХ(число; число_разрядов). Запишем необходимые условия согласно рис. 7.

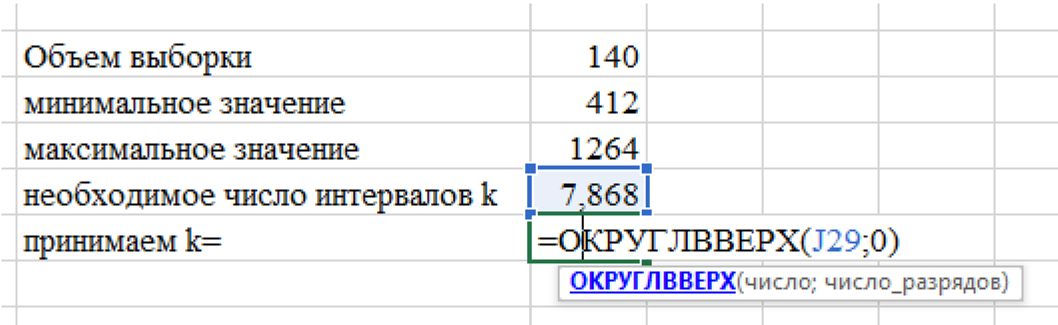


Рис. 7. Округление необходимого числа интервалов до целого числа

Создадим заголовки таблицы в ячейках M33–S33 согласно рис. 8.

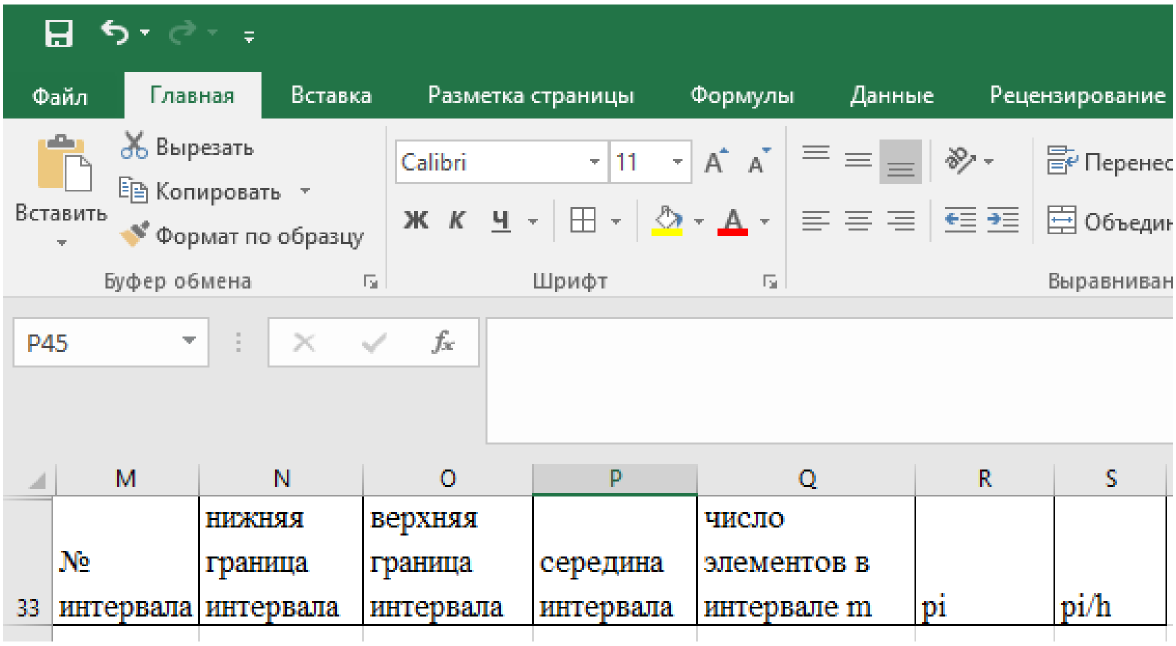


Рис. 8. Таблица для определения числа элементов в интервале

Длина интервала вычисляется по формуле

$$h = \frac{y_{\max} - y_{\min}}{n}, \qquad (32)$$

где $y_{\max}$ – максимальное значение выборки; $y_{\min}$ – минимальное значение выборки.

Для определения необходимой длины интервала запишем формулу в ячейку J31: «=(J28-J27)/J30». h = 106,5.

Первый интервал будет лежать в пределах $y_{\min} \ldots y_1$, где

$$y_1 = y_{\min} + h. \qquad (33)$$

Например, формула для вычисления верхней границы интервала 6 будет выглядеть следующим образом: «=N39+$J$31».

Результаты определения границ интервалов приведены на рис. 9.

	M	N	O	P	Q	R	S
31							
32							
33	№ интервала	нижняя граница интервала	верхняя граница интервала	середина интервала	число элементов в интервале m	pi	pi/h
34	1	412	518,5				
35	2	518,5	625				
36	3	625	731,5				
37	4	731,5	838				
38	5	838	944,5				
39	6	944,5	1051				
40	7	1051	1157,5				
41	8	1157,5	1264				

Рис. 9. Определение границ интервала

Для каждого интервала необходимо найти середину, используя формулу

$$h = \frac{y_{i-1} - y_i}{2}, \qquad (34)$$

где y_{i-1} – нижняя граница интервала; y_i – верхняя граница интервала.

Для определения числа элементов для каждого интервала используем функцию «СЧЕТЕСЛИМН» (рис. 10).

№ интервала	нижняя граница интервала	верхняя граница интервала	середина интервала	число элементов в интервале m	pi	pi/h
1	412	518,5	465,25	=СЧЁТЕСЛИМН(		
2	518,5	625	571,75			
3	625	731,5	678,25			
4	731,5	838	784,75			
5	838	944,5	891,25			
6	944,5	1051	997,75			
7	1051	1157,5	1104,25			
8	1157,5	1264	1210,75			

СЧЁТЕСЛИМН(**диапазон_условия1**; условие1; ...)

Рис. 10. Определение числа элементов, попадающих в *i*-й интервал

Для первого интервала выражение будет иметь вид:

«=СЧЕТЕСЛИМН($A$1:$F$24;">="&N34;$A$1:$F$24;"<="&O34)»,

где $A$1:$F$24 – исходные данные, среди которых ищем необходимые значения; «">="&N34» – условие для поиска значений среди исходных данных, которые должны быть больше или равны значению, указанному в ячейке N34 (нижней границы первого интервала); «<=»&O34 – условие для поиска значений среди исходных данных, значения которых меньше или равны значению, указанному в ячейке O34 (верхней границы первого интервала).

Значения, попадающие на границу интервала, должны быть учтены при расчетах только один раз, поэтому формула для расчета значений, попадающих во второй интервал, будет иметь вид:

«"=СЧЕТЕСЛИМН($A$1:$F$24;">"&N35;$A$1:$F$24;"<="&O35)"».

После выполнения расчетов сложим значения в столбце *Q*. Результат расчетов приведен на рис. 11.

Значение, полученное в ячейке Q42, должно совпадать с объемом выборки, вычисленным в ячейке J26.

№ интервала	нижняя граница интервала	верхняя граница интервала	середина интервала	число элементов в интервале m	pi	pi/h
1	412	518,5	465,25	18		
2	518,5	625	571,75	28		
3	625	731,5	678,25	23		
4	731,5	838	784,75	24		
5	838	944,5	891,25	24		
6	944,5	1051	997,75	9		
7	1051	1157,5	1104,25	8		
8	1157,5	1264	1210,75	6		
				140		

Рис. 11. Результаты расчетов по числу элементов для каждого интервала

Относительная частота (вероятность) попадания случайной величины в каждый интервал рассчитывается по формуле:

$$p_i = \frac{m_i}{n}, \tag{35}$$

$$\sum p_i = 1. \tag{36}$$

Например, для 5-го интервала формула будет иметь вид: «”=Q38/$Q$42”».

Результаты расчетов вероятности попадания случайной величины в каждый интервал приведены на рис. 12.

№ интервала	нижняя граница интервала	верхняя граница интервала	середина интервала	число элементов в интервале m	pi	pi/h
1	412	518,5	465,25	18	0,13	0,00121
2	518,5	625	571,75	28	0,2	0,00188
3	625	731,5	678,25	23	0,16	0,00154
4	731,5	838	784,75	24	0,17	0,00161
5	838	944,5	891,25	24	0,17	0,00161
6	944,5	1051	997,75	9	0,06	0,00060
7	1051	1157,5	1104,25	8	0,06	0,00054
8	1157,5	1264	1210,75	6	0,04	0,00040
				140	1	

Рис. 12. Результаты расчетов вероятности попадания случайной величины в каждый интервал

График, построенный по данным статистического ряда, называется *гистограммой*. Для его построения по оси абсцисс откладывают значения границ интервалов и на каждом из них (как на основании) строят прямоугольник, площадь которого должна быть равна относительной частоте, соответствующей данному интервалу. Следовательно, высота каждого прямоугольника равна $\frac{p_i}{h}$.

Для построения гистограммы используем функцию «Вставка – Диаграммы – Гистограмма – Гистограмма с группировкой» (рис. 13).

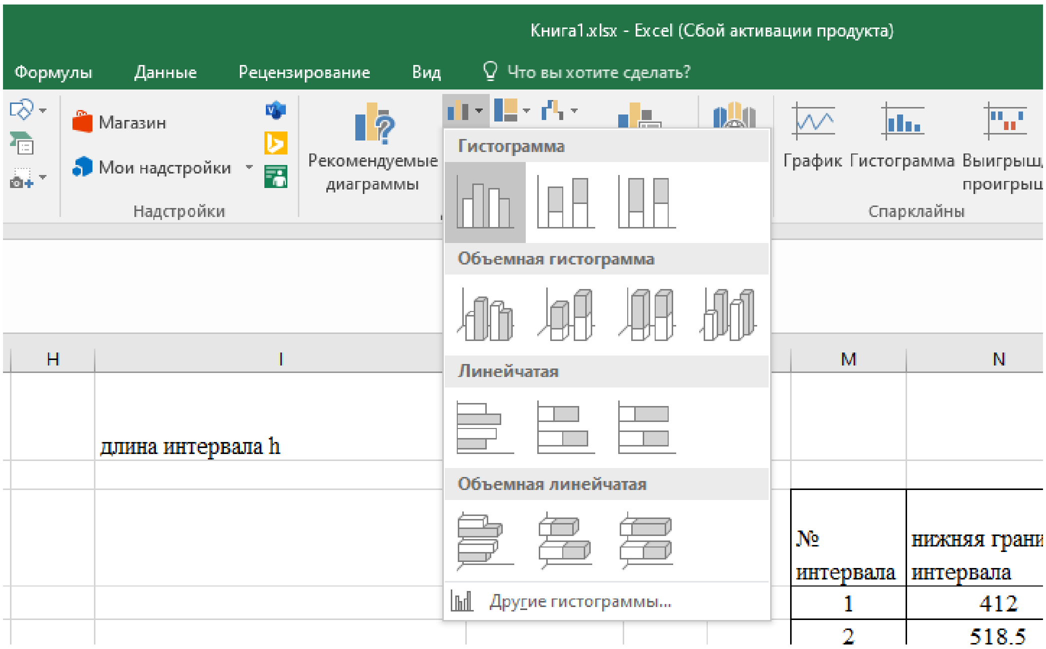


Рис. 13. Построение гистограммы

На выделенном пустом окне гистограммы щелкаем правой кнопкой мыши и выбираем «Выбрать данные».

В левом столбце «Элементы легенды (ряды)» нажимаем «Добавить» и в строку «Значения» указываем данные последнего столбца таблицы (рис. 14); подтверждаем выбор, нажав кнопку «ок».

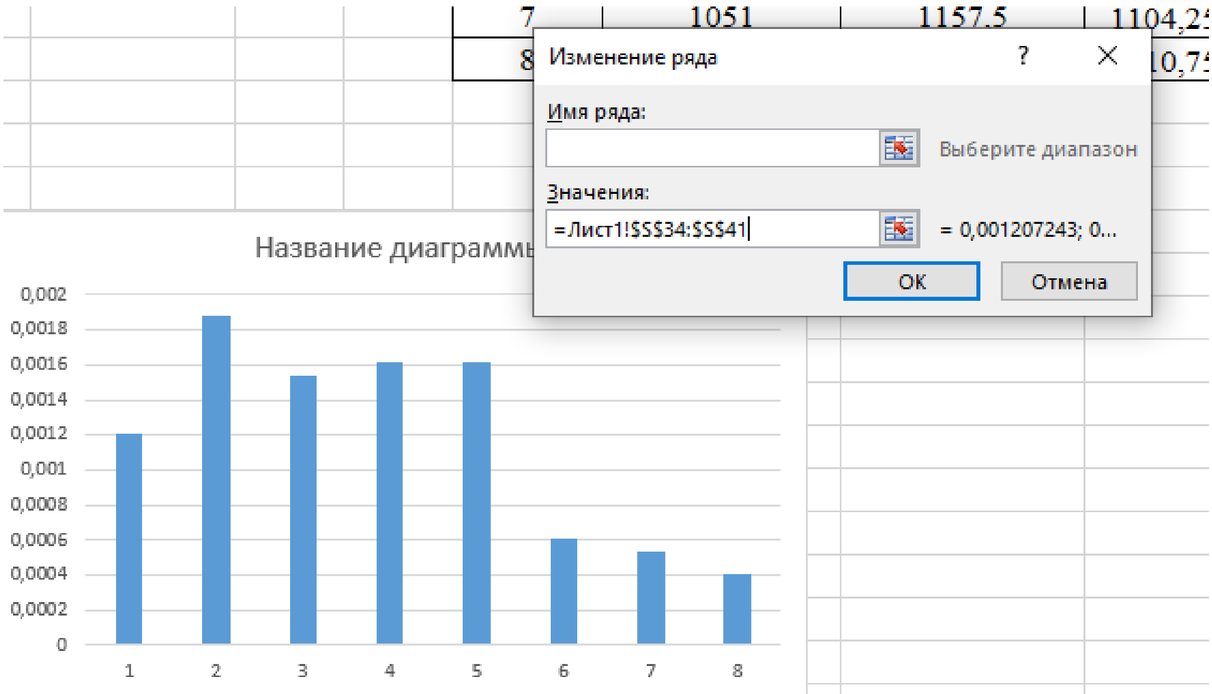


Рис. 14. Выбор исходных данных для построения гистограммы

В правом столбце «Подписи горизонтальной оси (категории)» нажимаем кнопку «Изменить» и выбираем данные из столбца «Середина интервала».

Результат построения гистограммы приведен на рис. 15.

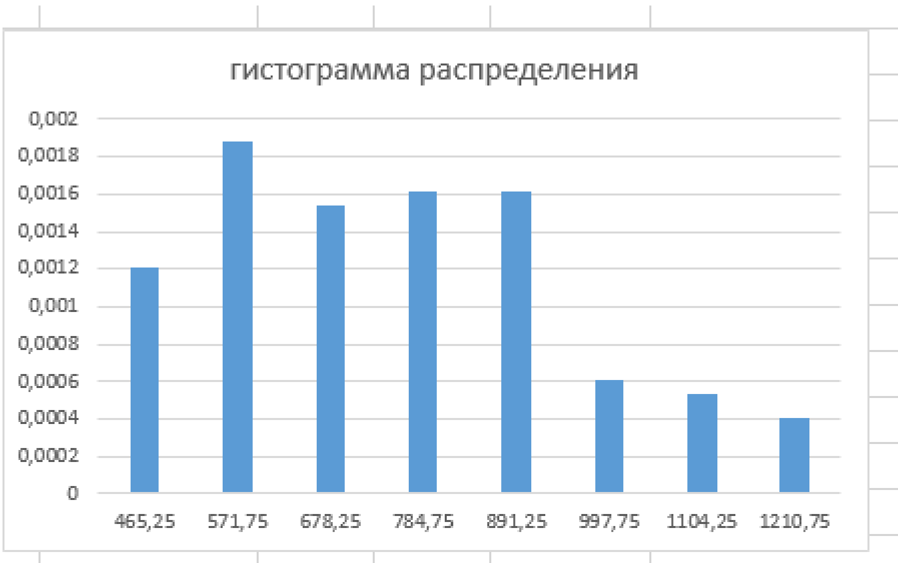


Рис. 15. Гистограмма распределения

Кривая распределения случайной величины представляет собой зависимость между ее значениями и соответствующими им вероятностями. В данном случае она строится по серединам интервалов y* и соответствующей относительной частоте p_i.

Для построения выбираем «Вставка – Диаграммы – Вставить график или диаграмму с областями – Линия» (рис. 16).

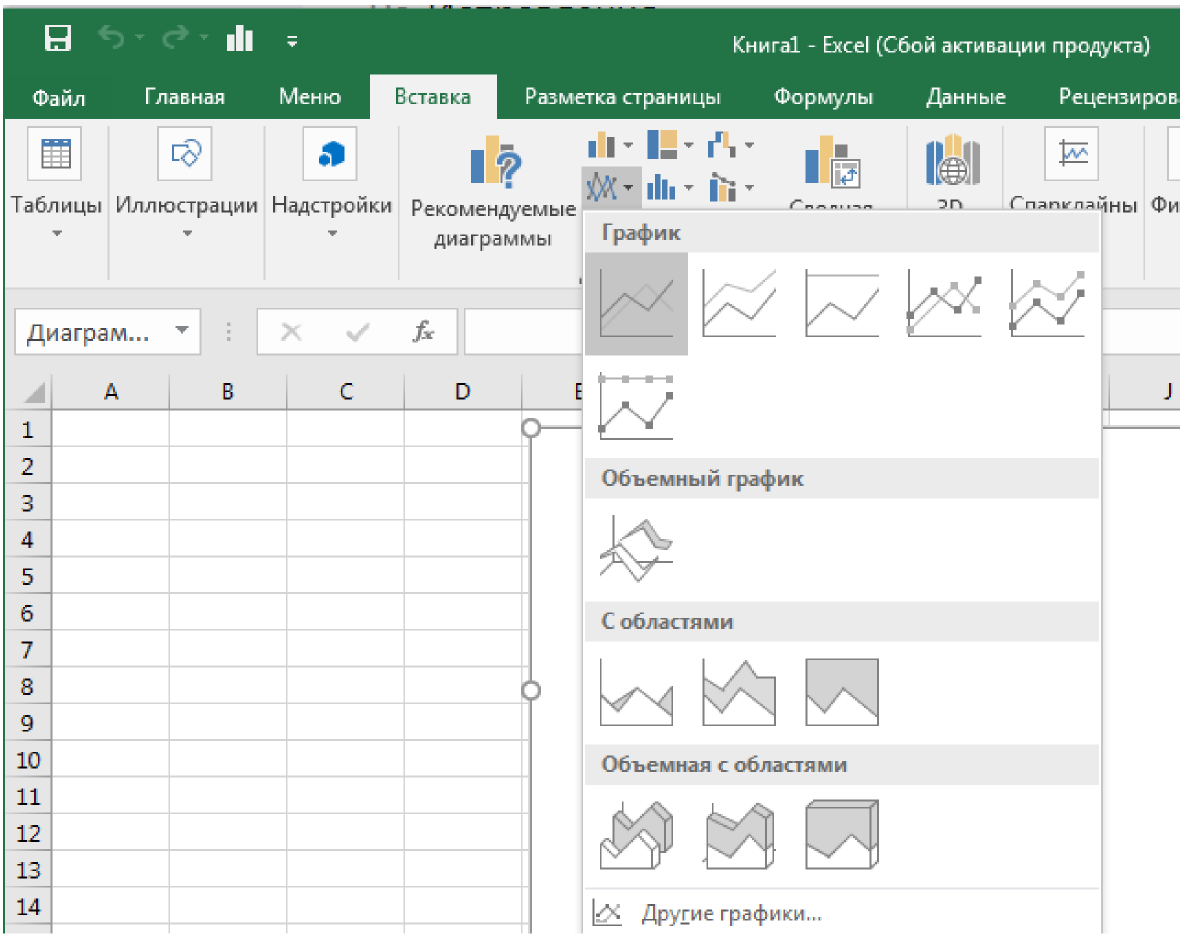
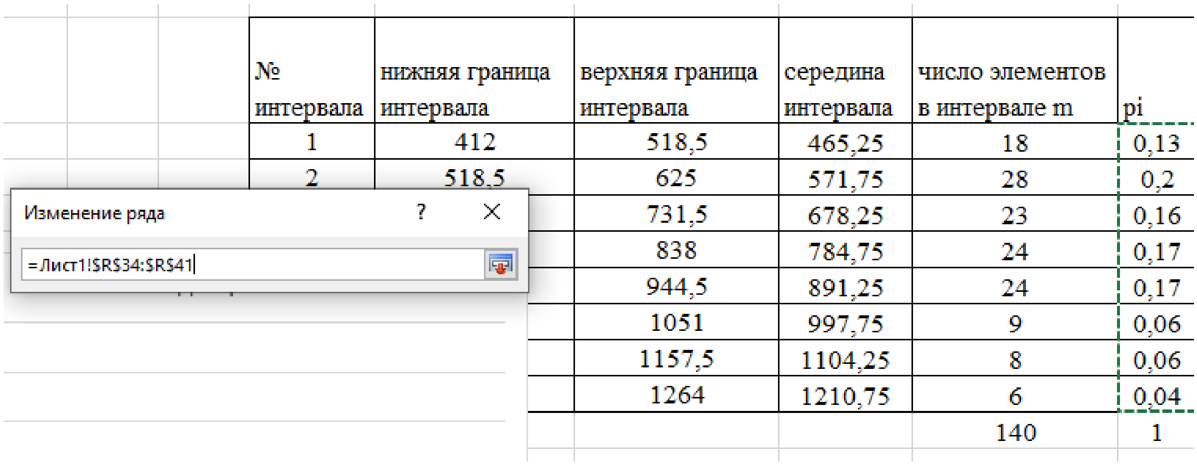


Рис. 16. Выбор типа графика для построения кривой распределения

После этого нажимаем на появившемся окне правой кнопкой мыши и выбираем команду «Выбрать данные…». В левом столбце «Элементы легенды (ряды)» нажимаем «Добавить» и в строке «Значения» указываем данные столбца *pi* таблицы (рис. 17); подтверждаем выбор, нажав кнопку «ок».

В правом столбце «Подписи горизонтальной оси (категории)» нажимаем кнопку «Изменить» и выбираем данные из столбца «Середина интервала».

Результат построения гистограммы приведен на рис. 18.



№ интервала	нижняя граница интервала	верхняя граница интервала	середина интервала	число элементов в интервале m	pi
1	412	518,5	465,25	18	0,13
2	518,5	625	571,75	28	0,2
		731,5	678,25	23	0,16
		838	784,75	24	0,17
		944,5	891,25	24	0,17
		1051	997,75	9	0,06
		1157,5	1104,25	8	0,06
		1264	1210,75	6	0,04
				140	1

Рис. 17. Выбор исходных данных для построения кривой распределения

Рис. 18. Кривая распределения шероховатости поверхности пиломатериалов в опыте № 25

Определение статистических показателей выборки

Для определения среднего арифметического выборки в ячейку J32 вводим «=СРЗНАЧ(A1:F24)».

Для расчета стандартного отклонения σ в ячейку J33 вводим функцию «=СТАНДОТКЛОН(A1:F24)».

Результаты расчетов приведены на рис. 19.

Объем выборки, n	140
минимальное значение	412
максимальное значение	1264
необходимое число интервалов k	7,9
принимаем k=	8
длина интервала h	106,5
выборочное среднее	749,7
стандартное отклонение	202,28

Рис. 19. Результаты расчета
среднего арифметического стандартного отклонения

Для расчета коэффициента вариации, средней квадратичной ошибки среднего значения, показателя точности среднего значения и ошибки среднего квадратического отклонения воспользуемся формулами (4), (5), (6), (7), введя их в ячейки J34–J37. Результаты расчетов приведены на рис. 20.

коэффициент вариации, %	0,27
средняя квадратичная ошибка среднего значения	0,02
показатель точности среднего значения, %	0,003
ошибка среднего квадратичного отклонения	12,09

Рис. 20. Результаты расчета основных показателей выборки

Чем меньше показатель точности, тем надежнее результаты исследования. Принято считать, что для лесной и деревообрабатывающей промышленности достаточная надежность будет обеспечена только в том случае, если показатель точности не превышает 5 %.

Выборочное среднее арифметическое представляет ценность, поскольку по нему можно судить об истинном среднем, генеральном среднем или математическом ожидании M_y. Представляет интерес отыскание величины максимальной ошибки Δ, которую допускают, предполагая, что $M_y = \bar{y}$. По этой причине требуется найти величину Δ, при которой выполняется условие:

$$\bar{y} - \Delta \le M_y \le \bar{y} + \Delta, \tag{37}$$

$$\Delta = \frac{t_{табл}\sigma}{\sqrt{n}}. \tag{38}$$

Однако нельзя абсолютно достоверно указать величину, т. к. расчет этой величины, как и любой статистический вывод, делают на основе результата эксперимента, а он заведомо содержит ошибки. Выводы, которые делают на основе неточных данных, принципиально не могут быть абсолютно достоверными, поэтому говорят о надежности статистического вывода, которую оценивают уровнем значимости (в деревообработке принимается равным 0,05). По уровню значимости и числу степеней свободы $f = n - 1$ определяют табличное значение критерия Стьюдента, используя табл. 7 прил. или вычисляя в *Excel* через функцию «СТЬЮДРАСПОБР(уровень значимости; число степеней свободы)».

Результат вычисления всех характеристик выборки приведен на рис. 21.

Объем выборки, n	140
минимальное значение	412
максимальное значение	1264
необходимое число интервалов k	7,9
принимаем k=	8
длина интервала h	106,5
выборочное среднее	749,7
стандартное отклонение	202,28
коэффициент вариации, %	0,27
средняя квадратичная ошибка среднего значения	0,02
показатель точности среднего значения, %	0,003
ошибка среднего квадратичного отклонения	12,09
табличное значение критерия Стьюента	1,98
математическое ожидание	33,80
доверительный интервал для математического ожидания	715,9мм <=Му<=783,5 мм

Рис. 21. Основные характеристики выборки

Для проверки на нормальность распределения нашей выборки введем дополнительные столбцы в ячейки T33–AA33 (рис. 22).

z1	z2	Φ(z1)	Φ(z2)	Pi*	Pi*n	(m-Pi*n)^2	(m-Pi*n)^2/Pi*n

Рис. 22. Дополнительные столбцы для проверки на нормальность распределения

Параметры z1 и z2 вычисляются по формуле (24).

Например, для 1 интервала

$$z_1 = \frac{450 - 749,7}{202,28} = -1,67; \qquad z_2 = \frac{518б5 - 749,7}{202,28} = -1,14.$$

Для вычисления величин $\Phi_0(z_1)$ и $\Phi_0(z_2)$ можно воспользоваться табл. 4 прил. или встроенной функцией *Excel* «$= НОРМСТРАСП(z_1) - 0,5$»: $p_i = \Phi_0(z_2) - \Phi_0(z_1) = -0,37 + 0,45 = 0,08$.

Найденные p_i заносим в столбец *X* нашей таблицы.

В столбец *Y* записываем значения $p_i n$ (*n* = 140 – число наблюдений).

В столбце *Z* приведены результаты вычислений величины $[m - p_i n]^2$.

В столбце AA приведены значения $\frac{[m - p_i n]^2}{p_i n}$. Данные расчетов приведены на рис. 23.

z1	z2	Φ(z1)	Φ(z2)	Pi*	Pi*n	(m-Pi*n)^2	(m-Pi*n)^2/Pi*n
-1,67	-1,14	-0,45	-0,37	0,08	11,06	48,18	4,36
-1,14	-0,62	-0,37	-0,23	0,14	19,91	65,39	3,28
-0,62	-0,09	-0,23	-0,04	0,20	27,35	18,91	0,69
-0,09	0,44	-0,04	0,17	0,20	28,65	21,60	0,75
0,44	0,96	0,17	0,33	0,16	22,89	1,24	0,05
0,96	1,49	0,33	0,43	0,10	13,95	24,47	1,75
1,49	2,02	0,43	0,48	0,05	6,48	2,31	0,36
2,02	2,54	0,48	0,49	0,02	2,30	13,72	5,97

Рис. 23. Результаты расчетов

Суммируя показания столбца АА, получаем значение

$$\chi^2 = \sum_{i=1}^{8} \frac{\left(m - p_i n\right)^2}{p_i n} = 4,36 + 3,28 + \ldots + 5,97 = 17,223.$$

Зададим уровень значимости q = 5 %. Зная число интервалов l, на которые разбиты опытные данные (у нас l = 5), вычислим число степеней свободы f = l – 3 = 5. По величинам q и f из табл. 6 прил. находим значение χ_q^2. В нашем случае $\chi_q^2 = 7,81$. Как видно, $\chi^2 > \chi_q^2$, поэтому гипотеза о нормальности данного распределения не может быть принята.

Проверка однородности двух дисперсий

Исходные данные для проверки однородности приведены в табл. 7.

Таблица 7

Результаты измерений двух выборок

№ выборки	Результаты наблюдений									
1	730	730	730	730	730	730	460	460	460	460
	460	460	460	460	460	460	460	460	460	460
	460	460	460	460	460	460	–	–	–	–
2	1100	1497	1230	829	1000	880	1050	740	712	730
	646	878	573	829	1132	1 013	926	730	740	720
	852 750	1220 710	1150 700	890 760	695 735	988 748	720 745	765 725	695 710	715 750

Проверку на однородность двух дисперсий делают при помощи критерия Фишера. Вначале определим для каждой выборки число степеней свободы и дисперсию.

Вычисление дисперсии происходит при помощи команды ДИСП, где в качестве аргумента указывается вся выборка. Для 1 выборки дисперсия равна 13458,46, а для второй – 38147,38.

В программу *Microsoft Excel* входит набор средств анализа данных (в виде надстройки «Пакет анализа»), предназначенный для решения задач обработки данных, математической статистики и ее приложений.

Для подключения пакета в *Microsoft Excel* 2016 следует нажать «Файл – Параметры» (рис. 24).

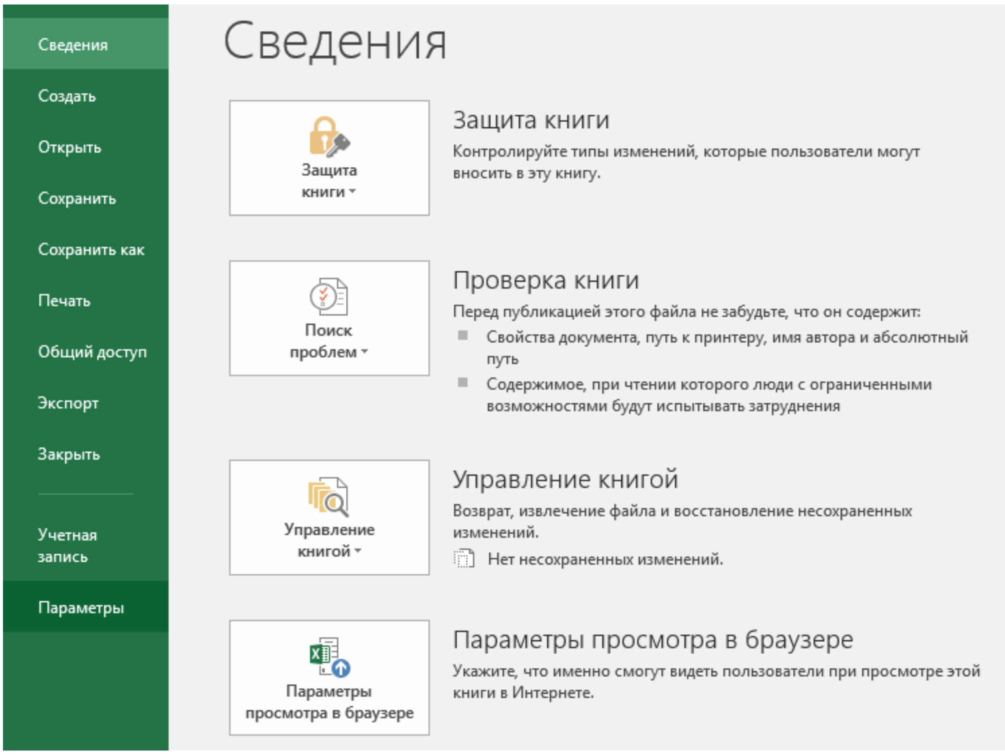


Рис. 24. Выбор вкладки «Параметры» в *Excel*

В появившемся окне «Параметры» в списке команд левой части окна выбрать команду «Надстройки», в списке «Управление» выбрать позицию «Надстройки Excel» и нажать кнопку «Перейти…» (рис. 25).

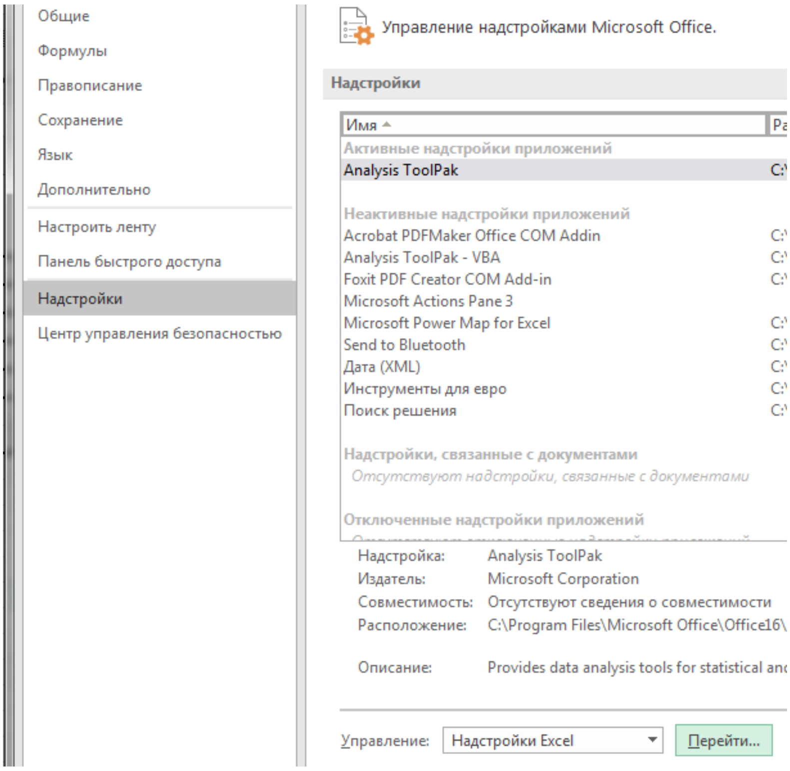


Рис. 25. Выбор команды «Надстройки *Excel*»

После этого появится окно «Надстройки», в котором надо установить флажок «Пакет анализа» и нажать кнопку «ок» (рис. 26).

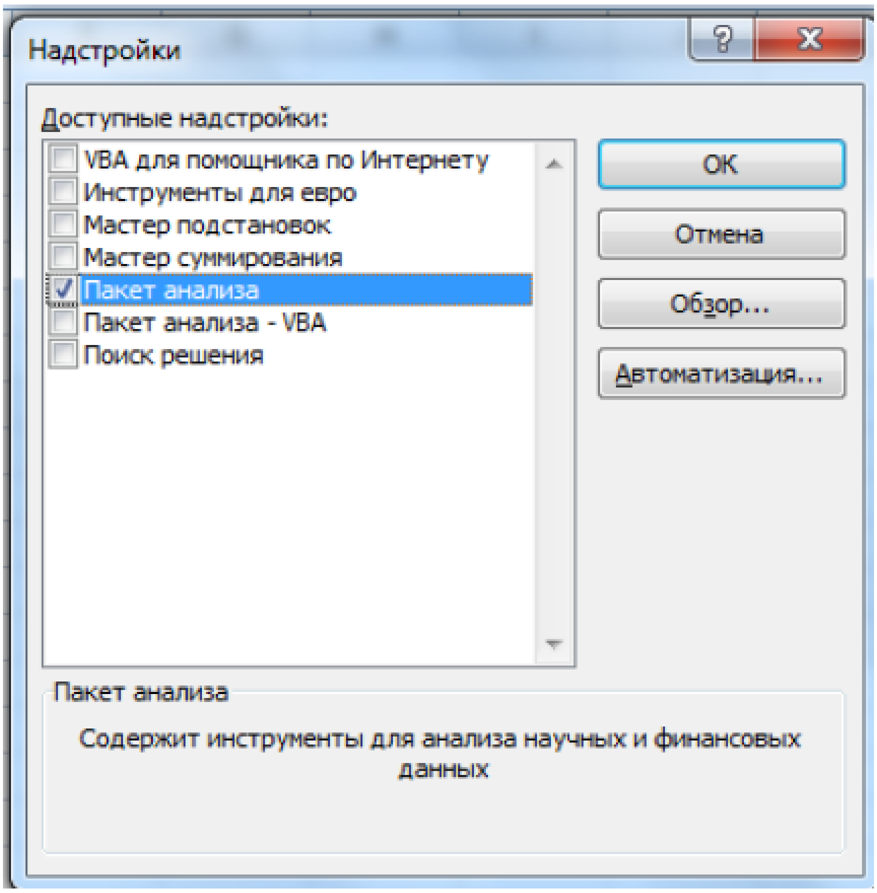


Рис. 26. Активация «Пакета анализа»

В результате подключения надстройки на вкладке «Данные» в группе «Анализ» станет доступна команда «Анализ данных». При выборе этой команды будет открываться окно «Анализ данных» (рис. 27).

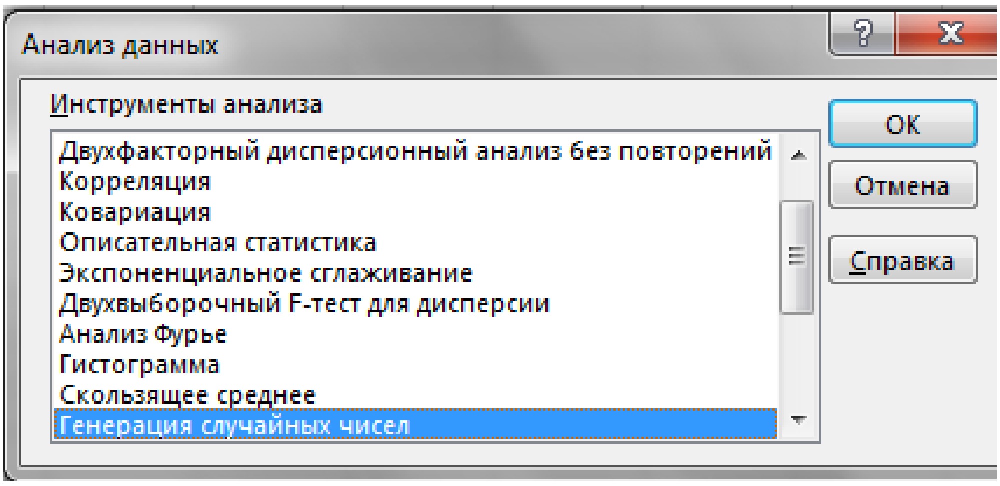


Рис. 27. Пакет «Анализ данных»

В пакете «Анализ данных» выбираем «Двухвыборочный F-тест для дисперсий». Заполненное диалоговое окно для наших данных приведено на рис. 28. Отметим, что средство требует, чтобы первым (в поле «Интервал переменной 1») задавался адрес диапазона ячеек, содержащих выборку с большей дисперсией.

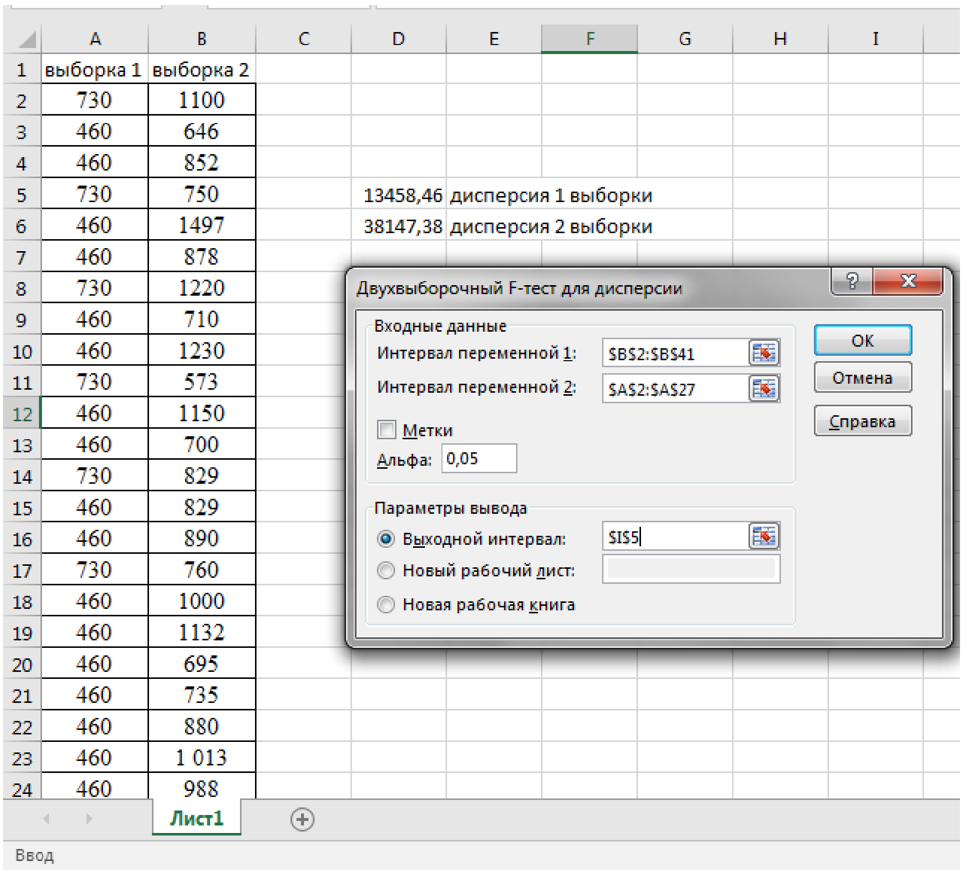


Рис. 28. Заполненные данные для вычисления двухвыборочного F-теста для дисперсий

В итоговой таблице приводятся следующие данные (рис. 29):

- среднее – выборочные средние арифметические выборок;
- дисперсия – несмещенные выборочные оценки дисперсий выборок;
- наблюдения – объемы выборок;
- корреляция Пирсона – выборочный коэффициент корреляции;
- гипотетическая разность средних – значение δ, которое задано в диалоговом окне;
- df – число степеней свободы (равны $n - 1$ и $m - 1$);
- F – расчетное значение критерия Фишера;
- F критическое одностороннее – табличное значение критерия Фишера (α, $n - 1$, $m - 1$).

Двухвыборочный F-тест для дисперсии		
	Переменная 1	Переменная 2
Среднее	849,45	522,3076923
Дисперсия	38147,38205	13458,46154
Наблюдения	40	26
df	39	25
F	2,834453399	
P(F<=f) одностороннее	0,003771687	
F критическое одностороннее	1,875535176	

Рис. 29. Значения двухвыборочного *F*-теста для дисперсии

Поскольку расчетное значение критерия Фишера больше табличного, то отвергаем гипотезу об однородности дисперсий.

Проверка на однородность нескольких дисперсий одинакового объема

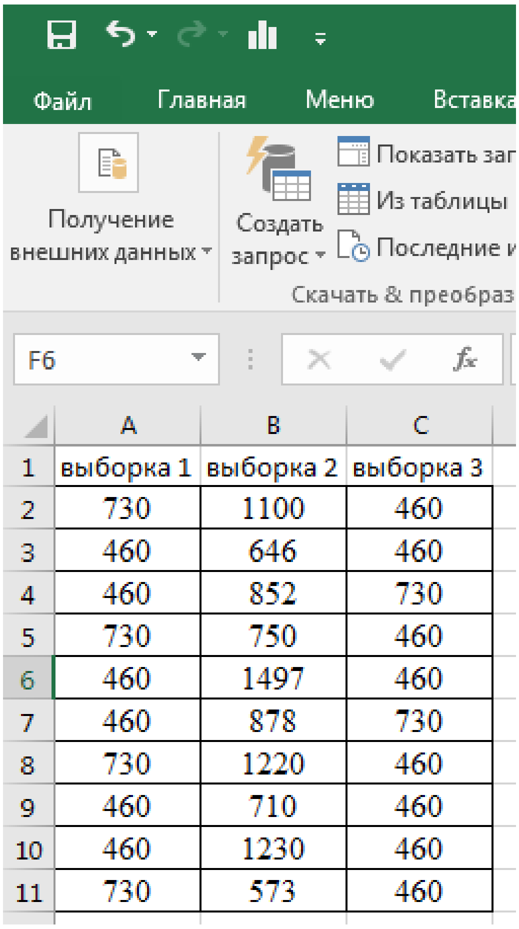
Использование *F*-критерия Фишера при числе более двух неэффективно, т. к. при этом в оценке участвуют только наибольшая и наименьшая дисперсии. Критерий Кохрена пригоден для случаев, когда число повторных опытов n во всех точках плана одинаково. Критерий Кохрена – это отношение максимальной дисперсии к сумме всех дисперсий:

$$G = \frac{\sigma_{\max}^2}{\sum_{i=1}^{m} \sigma_m^2}, \tag{39}$$

где m – количество выборочных дисперсий.

Затем по выбранному уровню значимости q, числу степеней свободы выборок $m_1 = n - 1$ и по количеству выборок m_2 из табл. 3 прил. находят величину $G_{таб}$. Если $G < G_{таб}$, то можно принять гипотезу об однородности дисперсий. В противном случае она отвергается.

Пусть у нас будут 3 выборки одинакового объема (рис. 30) $n = 10$.



	A	B	C
1	выборка 1	выборка 2	выборка 3
2	730	1100	460
3	460	646	460
4	460	852	730
5	730	750	460
6	460	1497	460
7	460	878	730
8	730	1220	460
9	460	710	460
10	460	1230	460
11	730	573	460

Рис. 30. Исходные данные для вычислений

При помощи команды ДИСП вычислим дисперсии для каждой выборки. При помощи команды МАКС определим, какая из выборок имеет максимальное значение ($\sigma^2_{\max} = 91083{,}156$), а также

$$\sum_{i=1}^{m} \sigma_m^2 = 19440 + 91083{,}156 + 12960 = 123483{,}16,$$

$$G = \frac{91083{,}156}{123483{,}16} = 0{,}74,$$

$$G_{таб} = 0{,}8.$$

Проверка однородности нескольких дисперсий, найденных по выборкам различного объема

Экспериментаторы часто планируют получение выборок одинакового объема, однако, если в опытах обнаруживаются промахи, то после их исключения объемы выборок оказываются различными. Пусть проверяется однородность некоторого числа *m* дисперсий: $s_1^2, s_2^2, s_3^2, \ldots, s_m^2$

. Теперь эти дисперсии найдены по выборкам различного объема – соответственно, $n_1, n_2, n_3, \ldots, n_m$. В этом случае используют критерий Бартлетта. Предварительно вычисляют величину s_y^2, представляющую собой среднее взвешенное значение дисперсий, взятое с учетом числа степеней свободы:

$$s_y^2 = \frac{\sum_{i=1}^{m} f_i s_i^2}{f}, \tag{40}$$

где $f = \sum_{i=1}^{m} f_i$ – числа степеней свободы соответствующих дисперсий:

$$f_i = n_i - 1. \tag{41}$$

Далее рассчитывают величину $B = V / C$. Здесь V и C равны:

$$\begin{gathered} V = 2,303\left(f \lg s_y^2 - \sum_{i=1}^{m} f_i \lg s_i^2 \right); \\ C = 1 + \frac{1}{3(m-1)}\left(\sum_{i=1}^{m} \frac{1}{f_i} - \frac{1}{f} \right). \end{gathered} \tag{42}$$

Затем из табл. 6 прил. при уровне значимости q и числе степеней свободы $k = m - 1$ отыскивают значение $\chi^2_{таб}$. Гипотеза об однородности дисперсий принимается, если $B \le \chi^2_{таб}$. В данной проверке требуется, чтобы объем каждой выборки был не менее четырех. Применение критерия Бартлетта является трудоемким. Кроме того, следует иметь в виду, что он весьма чувствителен к отклонениям от нормальности распределения.

Определение коэффициента корреляции

Во многих случаях целью экспериментальных исследований являются установление и изучение зависимости между некоторыми величинами. Если каждая из этих величин является случайной, то используют методы корреляционного анализа.

Будем говорить, что между двумя случайными величинами имеется статистическая связь, если при изменении одной из них меняется распределение другой. Для оценки статистической связи по данным эксперимента широко используется выборочный коэффициент корреляции. Пусть проведено n наблюдений и в каждом из них определялись значения двух параметров x и y. Найдем по двум выборкам среднее арифметическое $\overline{x}$ и $\overline{y}$, а также среднеквадратическое отклонение s_x, s_y. Выборочный коэффициент корреляции рассчитывается по формуле:

$$r_{xy} = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\left(n-1\right)s_x s_y} = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}}. \quad (43)$$

Коэффициент корреляции всегда лежит в пределах $-1 \le r_{xy} \le 1$. Он характеризует только линейную зависимость между случайными величинами.

Для определения значимости коэффициента корреляции рассчитывается экспериментальное значение *t*-критерия Стьюдента:

$$t_{расч} = \left|r_{xy}\right|\sqrt{\frac{n-2}{1-r_{xy}^2}}. \quad (44)$$

Расчетный критерий Стьюдента сравнивают с табличным значением *t*-критерия Стьюдента, найденным при выбранном уровне значимости q и числе степеней свободы $f = n - 2$. Если $t_{расч} \ge t_{таб}$, то между величинами x и y существует линейная статистическая связь.

Задача линейного регрессионного анализа состоит в восстановлении функциональной зависимости

$$y\left(x\right) \equiv M\left(Y / X = x\right) = a_0 + a_1 x \quad (45)$$

по результатам измерений

$$\left(x_1, y_1\right),\ \left(x_2, y_2\right),\ \ldots,\ \left(x_n, y_n\right). \quad (46)$$

Уравнение (эмпирическая регрессия):

$$\widehat{y} = \widehat{a}_0 + \widehat{a}_1 x \quad (47)$$

определяет прямую, которая является оценкой истинной линии регрессии. Необходимо вычислить точечные и интервальные оценки $\widehat{a}_0$, $\widehat{a}_1$ для параметров a_0, a_1 по результатам эксперимента и проверить значимость полученного уравнения регрессии [4–6].

Вычисление коэффициентов $\widehat{a}_0$, $\widehat{a}_1$ всегда производится с использованием метода наименьших квадратов, но этот метод фиксирует лишь «стратегию» получения эмпирических оценок, допуская различные «тактические приемы», что приводит к большому разнообразию конкретных математических постановок задач, методов и формул получения оценок $\widehat{a}_0$, $\widehat{a}_1$ даже в рассматриваемом здесь простейшем случае линейной регрессии. Отметим некоторые из них.

Коэффициенты регрессии можно вычислить:

а) минимизируя сумму квадратов отклонений:

$$E\left(\widehat{a}_0, \widehat{a}_1\right) = \sum_{i=1}^{n}\left(\widehat{a}_0 + \widehat{a}_1 x_i - y_i\right)^2; \tag{48}$$

б) численно решая систему уравнений:

$$\frac{\partial E\left(\widehat{a}_0, \widehat{a}_1\right)}{\partial \widehat{a}_0} = 0, \quad \frac{\partial E\left(\widehat{a}_0, \widehat{a}_1\right)}{\partial \widehat{a}_1} = 0; \tag{49}$$

в) решая (с использованием точных или итерационных методов) систему нормальных уравнений, предварительно сформировав ее в явном виде:

$$\begin{pmatrix} n & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} x_i^2 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \sum\limits_{i=1}^{n} y_i \\ \sum\limits_{i=1}^{n} x_i y_i \end{pmatrix}; \tag{50}$$

г) решая систему нормальных уравнений аналитически:

$$\widehat{a}_0 = \frac{\sum\limits_{i=1}^{n} y_i \sum\limits_{i=1}^{n}\left(x_i\right)^2 - \sum\limits_{i=1}^{n}\left(x_i y_i\right)\sum\limits_{i=1}^{n} x_i}{n\sum\limits_{i=1}^{n}\left(x_i\right)^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2}, \widehat{a}_1 = \frac{n\sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i}{n\sum\limits_{i=1}^{n}\left(x_i\right)^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} \tag{51}$$

или

$$\widehat{a}_1 = \frac{\sum\limits_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum\limits_{i=1}^{n}\left(x_i - \overline{x}\right)^2} = \frac{\sum\limits_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum\limits_{i=1}^{n} x_i^2 - n\overline{x}^2},$$

$$\widehat{a}_0 = \overline{y} - \widehat{a}_1\overline{x}, \ \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \ \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

или, если предварительно вычислены оценки дисперсий s_X^2, s_Y^2 и коэффициента корреляции $\widehat{\rho}$:

$$s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2, \; s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2, \tag{52}$$

$$r_{xy} = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}} = \frac{1}{(n-1)s_X s_Y}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right), \tag{53}$$

то

$$\widehat{a}_1 = r_{xy}\frac{s_Y}{s_X}, \; \widehat{a}_0 = \overline{y} - \widehat{a}_1\overline{x}.$$

Доверительные интервалы для коэффициентов регрессии $\widehat{a}_0$, $\widehat{a}_1$, соответствующие доверительной вероятности $p = 1 - \alpha$, имеют вид:

$$\begin{gathered}\widehat{a}_0 - t_{\alpha,n-2}\, s\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}} < a_0 < \widehat{a}_0 + t_{\alpha,n-2}\, s\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}}, \\ \widehat{a}_1 - \frac{t_{\alpha,n-2}\, s}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}} < a_1 < \widehat{a}_1 + \frac{t_{\alpha,n-2}\, s}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}}\end{gathered} \tag{54}$$

или

$$\begin{gathered}\widehat{a}_0 - t_{\alpha,n-2}\sqrt{\left(\frac{1}{n} + \frac{\overline{x}^2}{(n-1)s_X^2}\right)\frac{n-1}{n-2}s_Y^2\left(1 - \widehat{\rho}^2\right)} < a_0 <; \\ < \widehat{a}_0 + t_{\alpha,n-2}\sqrt{\left(\frac{1}{n} + \frac{\overline{x}^2}{(n-1)s_X^2}\right)\frac{n-1}{n-2}s_Y^2\left(1 - \widehat{\rho}^2\right)}; \\ \widehat{a}_1 - t_{\alpha,n-2}\frac{s_Y\sqrt{1 - \widehat{\rho}^2}}{s_X\sqrt{n-2}} < a_1 < \widehat{a}_1 + t_{\alpha,n-2}\frac{s_Y\sqrt{1 - \widehat{\rho}^2}}{s_X\sqrt{n-2}},\end{gathered} \tag{55}$$

где $t_{\alpha,n-2}$ – квантиль распределения Стьюдента, определяемый как корень уравнения

$$F_{n-2}\left(t_{\alpha,n-2}\right) = 1 - \alpha / 2,$$

где $F_{n-2}(t)$ – функция распределения Стьюдента с $n-2$ степенями свободы.

Доверительная область для всей линии регрессии определяется с помощью уравнений

$$y'(x) = \hat{a}_0 + \hat{a}_1 x - s\sqrt{2 f_{\alpha,2,n-2}} \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}},$$
$$y''(x) = \hat{a}_0 + \hat{a}_1 x + s\sqrt{2 f_{\alpha,2,n-2}} \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}, \qquad (56)$$

описывающих, соответственно, нижнюю и верхнюю границы области («полосы»), в которой с доверительной вероятностью $p = 1 - \alpha$ лежит истинная линия регрессии. Здесь $f_{\alpha,2,n-2}$ – квантиль распределения Фишера, определяемый как решение уравнения

$$F_{2,n-2}\left(f_{\alpha,2,n-2}\right) = 1 - \alpha; \qquad (57)$$

$F_{2,n-2}(x)$ – функция распределения Фишера с 2 и $n-2$ степенями свободы, s^2 – «остаточная» дисперсия, характеризующая рассеяние экспериментальных точек относительно линии регрессии:

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \qquad (58)$$

Для проверки значимости уравнения регрессии в целом используется критерий Фишера: если

$$\frac{s_Y^2}{s^2} > f_{\alpha,n-1,n-2}, \qquad (59)$$

то уравнение регрессии адекватно (статистически значимо) описывает результаты эксперимента при $(\alpha \cdot 100)$-процентном уровне значимости.

Отношение s_Y^2 / s^2 (полной и остаточной дисперсий) показывает, во сколько раз уравнение регрессии предсказывает результаты опыта лучше, чем среднее $\bar{y}$. Необходимо помнить, что доверительная оценка отклонения эмпирической линии регрессии от теоретической существенно ухудшается по мере удаления от среднего значения $\bar{x}$. В частности, по этой причине опасна экстраполяция эмпирической регрессионной зависимости за пределы интервала $(x_1,\ x_n)$, для которого она получена.

На этапе регрессионного анализа решаются следующие задачи:

1) выбор общего вида уравнения регрессии и определение параметров регрессии;

2) определение степени взаимосвязи результативного признака и факторов, проверка общего качества уравнения регрессии;

3) проверка статистической значимости каждого коэффициента уравнения регрессии и определение их доверительных интервалов.

Уравнение простой линейной регрессии имеет вид $\hat{y} = \hat{a}_0 + \hat{a}_1 x$. Параметры уравнения парной регрессии могут быть определены с помощью метода наименьших квадратов, который реализован в *Excel*. Для этого используется функция «Регрессия». Для ее вызова необходимо выбрать требуемое имя в окне диалога «Анализ данных». В результате появится следующее диалоговое окно (рис. 31).

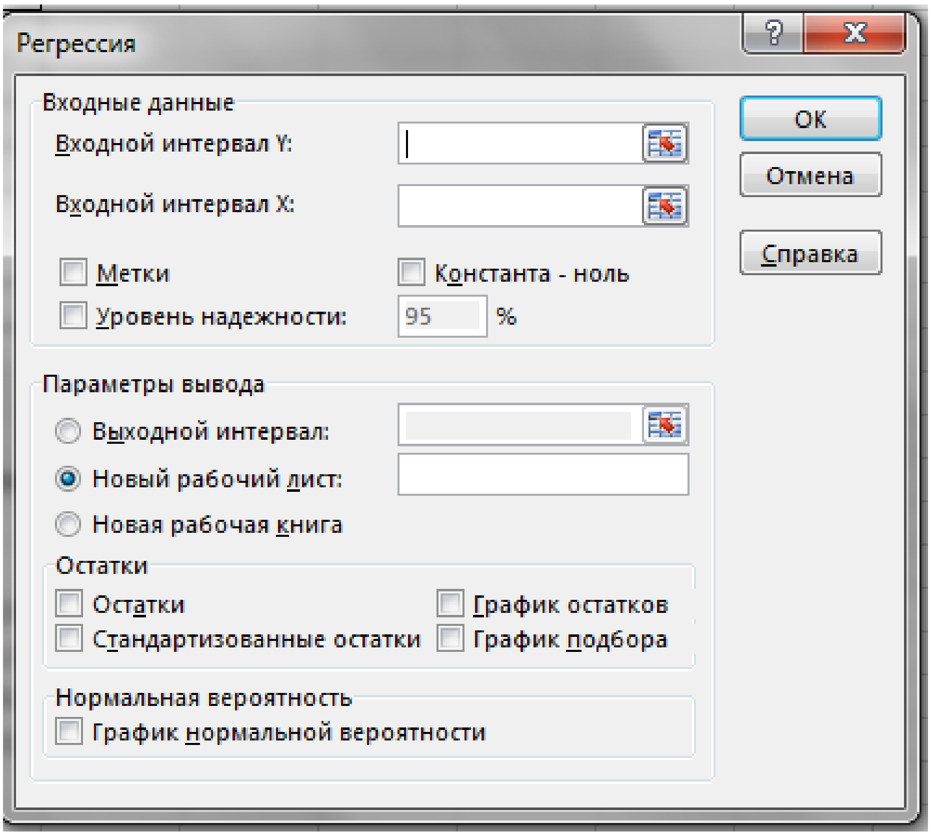


Рис. 31. Диалоговое окно «Регрессия»

В диалоговом окне задаются следующие параметры:

– *входной интервал Y* – вводится диапазон ячеек (один столбец), содержащих исходные данные по результирующему признаку;

– *входной интервал X* – вводится диапазон ячеек (число столбцов равно количеству признаков), содержащих исходные данные факторного признака;

– *метки* – флажок ставится, если первая строка содержит заголовок, в противном случае будут созданы стандартные заголовки автоматически;

– *уровень надежности* – флажок устанавливается, если требуется ввести значение уровня, отличное от 95 %. При выключенном флажке уровень надежности принимается равным 95 %;

– *константа-ноль* – флажок устанавливается в том случае, когда требуется, чтобы линия регрессии прошла через начало координат, т. е. $b = 0$;

– *параметры вывода* – указывается место, где будут указаны таблицы результатов анализа;

– *остатки* – при необходимости вывода столбца с остатками графическое представление остатков и подбора необходимо включить соответствующие флажки;

– *нормальная вероятность* – флажок устанавливается, если не требуется вывести график зависимости наблюдаемых значений от автоматически формируемых интервалов.

В результате работы инструмента «Регрессия» *Excel* формирует следующие данные (рис. 32).

11	ВЫВОД ИТОГОВ								
12									
13	*Регрессионная статистика*								
14	Множественный R	0,98454							
15	R-квадрат	0,96931							
16	Нормированный R-квадрат	0,95908							
17	Стандартная ошибка	1,72479							
18	Наблюдения	5							
19									
20	Дисперсионный анализ								
21		*df*	*SS*	*MS*	*F*	*ачимость F*			
22	Регрессия	1	281,875	281,875	94,7508	0,0023			
23	Остаток	3	8,92473	2,97491					
24	Итого	4	290,8						
25									
26		*Коэффициен*	*гартная о*	*татисти*	*-Значени*	*ижние 95*	*ерхние 95*	*жние 95,(*	*рхние 95,0%*
27	Y-пересечение	-1,06452	1,71083	-0,62222	0,57788	-6,50913	4,3801	-6,50913	4,3801
28	X	2,75269	0,28279	9,734	0,0023	1,85272	3,65266	1,85272	3,65266
29									
30									
31									
32	ВЫВОД ОСТАТКА					ВЫВОД ВЕРОЯТНОСТИ			
33									
34	*Наблюдение*	*дсказанн*	*Остатки*	*ртные остатки*		*Персентил*	*Y*		
35	1	4,44086	1,55914	1,0438		10	5		
36	2	12,6989	0,30108	0,20156		30	6		
37	3	23,7097	-0,70968	-0,47511		50	13		
38	4	7,19355	-2,19355	-1,46852		70	22		
39	5	20,957	1,04301	0,69827		90	23		

Рис. 32. Результат использования инструмента «Регрессия»

Для регрессионного анализа потребуется только часть данных, выделенных в таблицах желтым цветом.

Таблица «Регрессионная статистика» содержит линейный коэффициент корреляции R, коэффициент детерминации R^2, количество наблюдений n и другие параметры (рис. 33).

12		
13	*Регрессионная статистика*	
14	Множественный R	0,984535
15	R-квадрат	0,96931
16	Нормированный R-квадрат	0,95908
17	Стандартная ошибка	1,724793
18	Наблюдения	5

Рис. 33. Регрессионная статистика

Таблица «Дисперсионный анализ» содержит его параметры (табл. 8).

Таблица 8

Дисперсионный анализ

	df	*SS*	***MS***	***F***	***Значимость F***
Регрессия	1	281,88	**281,88**	**94,75**	**0,0023**
Остаток	3	8,92	**2,97**	–	–
Итого	4	290,8	–	–	–

Столбец *MS* – остаточная дисперсия $S^2_{ад}$ и общая дисперсия $S^2_{общ}$.

Столбец *F* – вычисленное значение критерия адекватности модели.

Столбец «Значимость *F*» оценивает адекватность полученной модели. Если значение меньше 0,05, то модель может считаться адекватной с вероятностью 0,95.

«Результативная таблица» (табл. 9) включает параметры a_0, a_1 уравнения регрессии, их статистические оценки, включая границы доверительных интервалов для коэффициентов уравнения регрессии.

Таблица 9

Результативная таблица

Обозначение коэффициента	Коэфф.	Стандартная ошибка	t-статистика	P-значение	Нижние 95 %	Верхние 95 %
Y-пересечение	**–1,06**	1,71	**–0,62**	0,58	**–6,51**	**4,38**
X	**2,75**	0,28	**9,73**	0,00	**1,85**	**3,65**

Таблица «Вывод остатка» (рис. 34) включает вычисленные по уравнению регрессии значения выходного параметра $\tilde{y}$ и значения остатков $y-\tilde{y}$.

ВЫВОД ОСТАТКА

Наблюдение	*Предсказанное Y*	*Остатки*	*Стандартные остатки*
1	4,44	1,56	1,04
2	12,70	0,30	0,20
3	23,71	-0,71	-0,48
4	7,19	-2,19	-1,47
5	20,96	1,04	0,70

Рис. 34. Вывод остатка

Между терминологией инструмента «Регрессия» и терминами, принятыми в математической статистике, имеется ряд расхождений. Согласование терминологии приводится в табл. 10.

Таблица 10

Основные показатели регрессионного анализа

Параметр инструмента «Регрессия»	**Статистический показатель**	**Обозначение**
Множественный *R*	Линейный коэффициент корреляции	r
R-квадрат	Коэффициент детерминации	r^2
Стандартная ошибка	Среднее квадратическое отклонение расчетных значений от фактических	σ
Наблюдения	Число наблюдений	n
F	Значение критерия Фишера	*F*
Значимость *F*	Оценка адекватности модели	–

Окончание табл. 10

MS	Дисперсия общая и остаточная	$S^2_{ад}, S^2_{общ}$
Y-пересечение	Свободный член регрессии	a_0
Переменная X1	Коэффициент регрессии	a_1
Предсказанное У	Расчетные значения результативного признака	y^*
Остатки	Отклонения расчетных значений от фактических	$y_i - y^*$

Средства надстройки «Анализ данных» имеют определенные ограничения и иногда удобнее пользоваться статистическими функциями или другими средствами *Excel*. Преимуществом функций перед данными средствами является то, что функции автоматически пересчитываются при любых изменениях, сделанных в выборке, тогда как средства надстройки «Анализ данных» необходимо выполнять заново, если выборка изменилась.

ПРИЛОЖЕНИЕ

Основные статистические таблицы

Таблица 1

Критические числа *mA*

№/№	Уровни значимости *q* %			№/№	Уровни значимости *q* %		
	1	5	10		1	5	10
1	–	–	–	46	13	15	16
2	–	–	–	47	14	16	17
3	–	–	–	48	14	16	17
4	–	–	–	49	15	17	18
5	–	–	0	50	15	17	18
6	–	0	0	51	15	18	19
7	–	0	0	52	16	18	19
8	0	0	1	53	16	18	20
9	0	1	1	54	17	19	20
10	0	1	1	55	17	19	20
11	0	1	2	56	17	20	21
12	1	2	2	57	18	20	21
13	1	2	3	58	18	21	22
14	1	2	3	59	19	21	22
15	2	3	3	60	19	21	23
16	2	3	4	61	20	22	23
17	2	4	4	62	20	22	24
18	3	4	5	63	20	23	24
19	3	4	5	64	21	23	24
20	3	5	5	65	21	24	25
21	4	5	6	66	22	24	25
22	4	5	6	67	22	25	26
23	4	6	7	68	22	25	26
24	5	6	7	69	23	25	27
25	5	7	7	70	23	26	27
26	6	7	8	71	24	26	28
27	6	7	8	72	24	27	28
28	6	8	9	73	25	27	28
29	7	8	9	74	25	28	29
30	7	9	10	75	25	28	29
31	7	9	10	76	26	28	30
32	8	9	10	77	26	29	30
33	8	10	11	78	27	29	31
34	9	10	11	79	27	30	31
35	9	11	12	80	28	30	32
36	9	11	12	81	28	31	32
37	10	12	13	82	28	31	33

Окончание табл. 1

№/№	Уровни значимости q %			№/№	Уровни значимости q %		
	1	5	10		1	5	10
38	10	12	13	83	29	32	33
39	11	12	13	84	29	32	33
40	11	13	14	85	30	32	34
41	11	13	14	86	30	33	34
42	12	14	15	87	31	33	35
43	12	14	15	88	31	34	35
44	13	15	16	89	31	34	36
45	13	25	16	90	32	35	36

Таблица 2

Значение $F_q(f_1 \cdot f_2)$, $q = 5$ %[1]

f_2	f_1							
	1	2	3	4	5	6	7	8
1	161	200	216	225	230	234	237	239
2	18,51	19,00	19,16	19,25	19,30	19,33	19,25	19,37
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73
8	5,32	4,76	4,07	3,84	3,69	3,58	3,50	3,44
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31

[1] f_1 – число степеней свободы числителя; f_2 – число степеней свободы знаменателя; q – уровень значимости.

Продолжение табл. 2

f_2	f_1							
	1	2	3	4	5	6	7	8
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27
31	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94

$$q = 5\ \%$$

f_2	f_1						
	9	10	12	15	20	30	∞
1	241	242	244	246	248	250	254
2	19,38	19,40	19,41	19,43	19,45	19,46	19,50
3	8,81	8,79	8,74	8,70	8,66	8,62	8,53
4	6,00	5,94	5,91	5,86	5,80	5,75	5,63
5	4,77	4,74	4,68	4,62	4,56	4,50	4,36
6	4,10	4,06	4,00	3,94	3,87	3,81	3,67
7	3,68	3,64	3,57	3,51	3,44	3,38	3,23
8	3,39	3,35	3,28	3,22	3,15	3,08	2,93
9	3,18	3,14	3,07	3,01	2,94	2,86	2,71
10	3,02	2,98	2,91	2,85	2,77	2,70	2,54
11	2,90	2,85	2,79	2,72	2,65	2,57	2,40
12	2,80	2,75	2,69	2,62	2,54	2,47	2,30
13	2,71	2,67	2,60	2,53	2,46	2,38	2,21
14	2,65	2,60	2,53	2,46	2,39	2,31	2,13
15	2,59	2,54	2,48	2,40	2,33	2,25	2,07
16	2,54	2,49	2,42	2,35	2,28	2,19	2,01
17	2,49	2,45	2,38	2,31	2,23	2,15	1,96
18	2,46	2,41	2,34	2,27	2,19	2,11	1,92
19	2,42	2,38	2,31	2,23	2,16	2,07	1,88
20	2,39	2,35	2,28	2,20	2,12	2,04	1,84
21	2,37	2,32	2,25	2,18	2,10	2,01	1,81
22	2,34	2,30	2,23	2,15	2,07	1,98	1,78
23	2,32	2,27	2,20	2,13	2,05	1,96	1,76
24	2,30	2,25	2,18	2,11	2,03	1,94	1,73
25	2,28	2,24	2,16	2,09	2,01	1,92	1,71
26	2,27	2,22	2,15	2,07	1,99	1,90	1,69
27	2,25	2,20	2,13	2,06	1,97	1,88	1,67
28	2,24	2,19	2,12	2,04	1,96	1,87	1,65
29	2,22	2,18	2,10	2,03	1,94	1,85	1,64
30	2,21	2,16	2,09	2,01	1,93	1,84	1,62
40	2,12	2,08	2,00	1,92	1,84	1,74	1,51

Продолжение табл. 2

f_2	f_1						
	9	10	12	15	20	30	∞
60	2,04	1,99	1,92	1,84	1,75	1,65	1,39
120	1,96	1,91	1,83	1,75	1,66	1,55	1,25
∞	1,88	1,83	1,75	1,67	1,57	1,46	1,00

$q = 1\ \%$

f_2	f_1							
	1	2	3	4	5	6	7	8
1	4052	4999	5403	5625	5764	5859	5928	5982
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37
3	34,12	30,82	29,46	28,71	28,42	27,91	27,67	27,49
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,61
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17
40	7,31	5,18	4,31	3,85	3,51	3,29	3,12	2,99
60	7,08	4,98	4,13	3,63	3,34	3,12	2,95	2,82
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51

Окончание табл. 2

$q = 1\ \%$

f_2	f_1						
	9	10	12	15	20	30	∞
1	6022	6056	6106	6157	6209	6261	6366
2	99,39	99,40	99,42	99,43	99,45	99,47	99,50
3	27,35	27,23	27,05	26,87	26,69	26,50	26,13
4	14,66	14,55	14,37	14,20	14,02	13,84	13,46
5	10,16	10,05	9,89	9,72	9,55	9,38	9,02
6	7,98	7,87	7,72	7,56	7,40	7,23	6,88
7	6,72	6,62	6,47	6,31	6,16	5,99	5,65
8	5,91	5,81	5,67	5,52	5,36	5,20	4,86
9	5,35	5,26	5,11	4,96	4,81	4,65	4,31
10	4,94	4,85	4,71	4,56	4,41	4,25	3,94
11	4,63	4,54	4,40	4,25	4,10	3,94	3,60
12	4,39	4,30	4,16	4,01	3,86	3,70	3,36
13	4,19	4,10	3,96	3,82	3,66	3,51	3,17
14	4,03	3,94	3,80	3,66	3,51	3,35	3,00
15	3,89	3,80	2,67	3,52	3,37	3,21	2,87
16	3,78	3,69	3,55	3,41	3,26	3,10	2,75
17	3,68	3,59	3,46	3,31	3,16	3,00	2,65
18	3,60	3,51	3,37	3,23	3,08	2,92	2,57
19	3,52	3,43	3,30	3,15	3,00	2,84	2,49
20	3,46	3,37	3,23	3,09	2,94	2,78	2,42
21	3,40	3,31	3,17	3,03	2,88	2,72	2,36
22	3,35	2,26	3,12	2,98	2,83	2,67	2,31
23	3,30	3,21	3,07	2,93	2,78	2,62	2,26
24	3,26	3,17	3,03	2,89	2,74	2,58	2,21
25	3,22	3,13	2,99	2,85	2,70	2,54	2,17
26	3,18	3,09	2,96	2,81	2,66	2,50	2,13
27	3,15	3,06	2,93	2,78	2,63	2,47	2,10
28	3,12	3,03	2,90	2,75	2,60	2,44	2,06
29	3,09	3,00	2,87	2,73	2,57	2,41	2,03
30	3,07	2,98	2,84	2,70	2,55	2,39	2,01
40	2,89	2,80	2,66	2,52	2,37	2,20	1,80
60	2,72	2,63	2,50	2,35	2,20	2,03	1,60
120	2,56	2,47	2,34	2,19	2,03	1,86	1,38
∞	2,41	2,32	2,18	2,04	1,88	1,70	1,00

Таблица 3

Значение $G_q(f_1 K)$, q = 5 %[2]

K	f													
	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0,99	0,98	0,94	0,91	0,88	0,85	0,83	0,82	0,80	0,79	0,73	0,66	0,58	0,50
3	0,97	0,87	0,80	0,75	0,71	0,68	0,65	0,63	0,62	0,60	0,55	0,47	0,40	0,33
4	0,91	0,77	0,68	0,63	0,59	0,56	0,54	0,52	0,50	0,49	0,44	0,37	0,31	0,25
5	0,84	0,68	0,60	0,54	0,51	0,48	0,46	0,44	0,42	0,41	0,36	0,31	0,25	0,20
6	0,78	0,62	0,53	0,48	0,44	0,42	0,40	0,38	0,37	0,36	0,31	0,26	0,21	0,17
7	0,73	0,56	0,48	0,43	0,40	0,37	0,35	0,34	0,33	0,32	0,28	0,23	0,18	0,14
8	0,68	0,52	0,44	0,39	0,36	0,34	0,32	0,30	0,29	0,28	0,25	0,20	0,16	0,13
9	0,64	0,48	0,40	0,36	0,33	0,31	0,29	0,28	0,27	0,26	0,22	0,18	0,14	0,11
10	0,60	0,45	0,37	0,33	0,30	0,28	0,27	0,25	0,24	0,24	0,20	0,17	0,13	0,10
12	0,54	0,39	0,33	0,29	0,26	0,24	0,23	0,22	0,21	0,20	0,17	0,14	0,11	0,08
15	0,47	0,33	0,28	0,24	0,22	0,20	0,19	0,18	0,17	0,17	0,14	0,11	0,09	0,07
20	0,39	0,27	0,22	0,19	0,17	0,16	0,15	0,14	0,14	0,13	0,11	0,09	0,07	0,05
24	0,34	0,24	0,19	0,17	0,15	0,14	0,13	0,12	0,12	0,11	0,09	0,07	0,06	0,04
30	0,29	0,20	0,16	0,14	0,12	0,11	0,11	0,10	0,10	0,09	0,08	0,06	0,05	0,03
40	0,24	0,16	0,13	0,11	0,10	0,09	0,08	0,08	0,07	0,07	0,06	0,05	0,03	0,03
60	0,17	0,11	0,09	0,08	0,07	0,06	0,06	0,06	0,05	0,05	0,04	0,03	0,02	0,02
120	0,10	0,06	0,05	0,04	0,04	0,03	0,03	0,03	0,03	0,03	0,02	0,02	0,01	0,01
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

q = 1 %

K	f													
	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0,99	0,99	0,98	0,96	0,94	0,92	0,90	0,88	0,87	0,85	0,79	0,71	0,61	0,50
3	0,99	0,94	0,88	0,83	0,79	0,76	0,73	0,71	0,69	0,67	0,61	0,52	0,42	0,33
4	0,97	0,86	0,78	0,72	0,68	0,64	0,61	0,59	0,57	0,55	0,49	0,41	0,33	0,25
5	0,93	0,79	0,70	0,63	0,59	0,55	0,53	0,50	0,49	0,47	0,41	0,34	0,26	0,20
6	0,88	0,72	0,63	0,56	0,52	0,49	0,46	0,44	0,42	0,41	0,35	0,29	0,22	0,17
7	0,84	0,66	0,57	0,51	0,47	0,43	0,41	0,39	0,38	0,36	0,31	0,25	0,19	0,14
8	0,79	0,62	0,52	0,46	0,42	0,39	0,37	0,35	0,34	0,32	0,28	0,22	0,17	0,13
9	0,75	0,57	0,48	0,43	0,39	0,36	0,34	0,32	0,31	0,30	0,25	0,20	0,15	0,11
10	0,72	0,54	0,45	0,39	0,36	0,33	0,31	0,29	0,28	0,27	0,23	0,18	0,14	0,10
12	0,65	0,48	0,39	0,33	0,31	0,29	0,27	0,25	0,24	0,23	0,20	0,15	0,12	0,08
15	0,57	0,41	0,33	0,29	0,26	0,24	0,22	0,21	0,20	0,19	0,16	0,13	0,09	0,07
20	0,48	0,33	0,27	0,23	0,20	0,19	0,17	0,16	0,16	0,15	0,12	0,10	0,07	0,05
24	0,42	0,29	0,23	0,20	0,18	0,16	0,15	0,14	0,13	0,13	0,11	0,08	0,06	0,04
30	0,36	0,24	0,19	0,16	0,15	0,13	0,12	0,12	0,11	0,11	0,09	0,07	0,05	0,03
40	0,29	0,19	0,15	0,13	0,11	0,10	0,10	0,09	0,09	0,08	0,07	0,05	0,04	0,03
60	0,22	0,14	0,11	0,09	0,08	0,07	0,07	0,06	0,06	0,06	0,05	0,03	0,02	0,02
120	0,12	0,08	0,06	0,05	0,04	0,04	0,04	0,03	0,03	0,03	0,02	0,02	0,01	0,01
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

[2] f – число степеней свободы; K – число выборок; q – уровень значимости.

Таблица 4

Нормированная функция Лапласа*

$$\Phi_0(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{\nu}{2}} d\nu$$

z	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	40	80	120	160	199	239	279	319	359
0,1	398	438	478	517	557	596	636	675	714	753
0,2	793	832	871	910	948	987	*026	*064	*103	*141
0,3	0,1179	217	255	293	331	368	406	443	480	517
0,4	554	591	628	664	700	736	772	808	844	879
0,5	915	950	985	*019	*054	*088	*123	*157	*190	*224
0,6	0,2257	291	324	357	389	422	454	486	517	549
0,7	580	611	642	673	703	734	764	794	823	852
0,8	881	910	939	967	995	*023	*051	*078	*106	*133
0,9	0,3159	186	212	238	264	289	315	340	365	389
1,0	413	437	461	485	508	583	554	577	599	621
1,1	643	665	686	708	729	749	770	790	810	830
1,2	649	869	868	907	925	944	962	980	997	*015
1,3	0,4032	49	66	82	99	115	131	147	162	177
1,4	192	207	222	236	251	265	279	292	306	319
1,5	332	345	357	370	382	394	406	418	429	441
1,6	452	463	474	484	495	505	515	525	535	545
1,7	554	564	573	582	591	599	608	616	625	633
1,8	641	649	656	664	671	678	686	693	699	706
1,9	713	719	726	732	738	744	750	756	761	767
2,0	772	778	783	788	793	798	803	808	812	817
2,1	821	826	830	834	838	842	846	850	854	857
2,2	860	864	867	871	874	877	880	883	886	889
2,3	892	895	898	900	903	906	908	911	913	915
2,4	918	920	922	924	926	928	930	932	934	936
2,5	937	939	941	942	944	946	947	949	950	952
2,6	953	954	956	957	958	959	960	962	963	964
2,7	965	966	967	968	969	970	971	971	972	973
2,8	0,4974	975	975	976	977	978	978	979	980	980
2,9	981	981	982	983	983	984	984	985	985	986
3,0	986	986	987	987	988	988	988	989	989	989
3,1	990	990	990	991	991	991	992	992	992	992
3,2	993	993	993	993	994	994	994	994	994	994

Окончание табл. 4

z	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
3,3	995	995	995	995	995	995	996	996	996	996
3,4	996	996	996	996	997	997	997	997	997	997
3,5	997	997	997	997	998	998	988	998	998	998
3,6	998	998	998	998	998	998	998	998	998	998
3,7	998	998	999	999	999	999	999	999	999	999
3,8	999	999	999	999	999	999	999	999	999	999
3,9	999	999	999	999	999	999	999	999	999	999
4,0	999	999	999	999	999	999	999	999	999	999
4,1	999	999	999	999	999	999	999	999	999	999
4,2	999	999	999	999	999	999	999	999	999	999
4,3	999	999	999	999	999	999	999	999	999	999
4,4	999	999	999	999	999	999	999	999	999	999
4,5	999	999	999	999	999	999	999	999	999	999
4,6	999	999	999	999	999	999	999	999	999	999
4,7	999	999	999	999	999	999	999	999	999	999
4,8	999	999	999	999	999	999	999	999	999	999
4,9	999	999	999	999	999	999	999	999	999	999
5,0	999	999	999	999	999	999	999	999	999	999

**Примечание*: Если перед последними тремя десятичными знаками стоит «*», то это означает, что первый десятичный знак надо смотреть в графе «0» следующей строки.

Таблица 5

Критерий λ Колмогорова-Смирнова для сопоставления эмпирического распределения с теоретическим

λ	P(λ)	λ	P(λ)	λ	P(λ)
0,0	1,000	0,7	0,711	1,4	0,040
0,1	1,000	0,8	0,544	1,5	0,022
0,2	1,000	0,9	0,393	1,6	0,012
0,3	1,000	1,0	0,270	1,7	0,006
0,4	0,997	1,1	0,178	1,8	0,003
0,5	0,964	1,2	0,112	1,9	0,002
0,6	0,864	1,3	0,068	2,0	0,001

Таблица 6

Значение $\chi_q^2(f)$[3]

f	q %		f	q %		f	q %		f	q %	
	5	1		5	1		5	1		5	1
1	3,84	6,63	14	23,7	29,1	27	40,1	47,0			
2	5,99	9,21	15	25,0	30,6	28	41,3	48,3	40	55,8	63,7
3	7,81	11,3	16	26,3	32,0	29	42,6	49,6	41	56,9	65,0
4	9,49	13,3	17	27,6	33,4	30	43,8	50,9	42	58,1	66,2
5	11,1	15,1	18	28,9	34,8	31	45,0	52,2	43	59,3	67,5
6	12,6	16,8	19	30,1	36,2	32	46,2	53,5	44	60,5	68,7
7	14,1	18,5	20	31,4	37,6	33	47,4	54,8	45	61,7	70,0
8	15,5	20,1	21	32,7	38,9	34	48,6	56,1	46	62,8	71,2
9	16,9	21,7	22	33,9	40,3	35	49,8	57,3	47	64,0	72,4
10	18,3	23,2	23	35,2	41,6	36	51,0	58,6	48	65,2	73,7
11	19,7	24,7	24	36,4	43,0	37	52,2	59,9	49	66,3	74,9
12	21,0	26,2	25	37,7	44,3	38	53,4	61,2	50	67,5	76,2
13	22,4	27,7	26	38,9	45,6	39	54,6	62,4			

Таблица 7

Значение t_{qf}[4]

f	q %		f	q %		f	q %	
	5	1		5	1		5	1
1	12,71	63,66	14	2,15	2,98	27	2,05	2,77
2	4,30	9,93	15	2,13	2,95	28	2,05	2,76
3	3,18	5,84	16	2,12	2,92	29	2,04	2,76
4	2,78	4,60	17	2,11	2.90	30	2,04	2,75
5	2,57	4,03	18	2,10	2,88	49	2,02	2,70
6	2,45	3,71	19	2,09	2,86	50	2,01	2,68
7	2,37	3,50	20	2,09	2,85	60	2,00	2,66
8	2,31	3,36	21	2,08	2,83	80	1,99	2,64
9	2,26	3,25	22	2,07	2,82	100	1,98	2,63
10	2,23	3,17	23	2,07	2,81	120	1,98	2,62
11	2,20	3,11	24	2,06	2,80	200	1,97	2,60
12	2,18	3,06	25	2,06	2,79	500	1,96	2,59
13	2,16	3,01	26	2,06	2,78	∞	1,96	2,58

[3] f – число степеней свободы; q – уровень значимости.

[4] f – число степеней свободы; q – уровень значимости.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Смирнов М. В., Дунин-Барковский И. В. Краткий курс математической статистики для технических приложений. Наука., 1969. 512 с.
2. Длин А. М. Математическая статистика в технике. 3-е изд., перераб. М. : Советская наука, 1958. 466 с.
3. Таблицы математической статистики. 3-е изд. М. : Наука, 1983. 416 с.
4. Демин С. Е., Демина Е. Л. Математическая статистика : учеб.-метод. пособие. Нижний Тагил : НТИ (филиал) УрФУ, 2016. 284 с.
5. Методы и средства научных исследований. Методы планирования и обработки результатов экспериментов: учебное пособие / А. Н. Чубинский, Д. С. Русаков, И. М. Батырева [и др.]. СПб : СПбГЛТУ, 2018. 107 с.
6. Шалабанов А. К., Роганов Д. А. Практикум по эконометрике с применением MS EXCEL. Линейные модели парной и множественной регрессии. Казань : Академия управления ТИСБИ, 2008. 198 с.

ОГЛАВЛЕНИЕ

Учебное издание

***Мялицин** Алексей Владимирович*

Методика планирования экспериментов и обработки их результатов при исследовании технологических процессов в деревообрабатывающей промышленности

Редактор Н. Ф. Тофан
Оператор компьютерной верстки О. А. Казанцева

ISBN 978-5-94984-877-7

9 785949 848777

Подписано в печать 05.09.2023. Формат 60х84/16.
Бумага офсетная. Цифровая печать.
Уч.-изд. л. 3,92. Печ. л. 4,42.
Тираж 300 экз. (1-й завод 36 экз.).
Заказ № 7712

ФГБОУ ВО «Уральский государственный лесотехнический университет».
620100, Екатеринбург, Сибирский тракт, 37.
Редакционно-издательский отдел. Тел.: 8(343) 221-21-44.

Типография ООО «ИЗДАТЕЛЬСТВО УЧЕБНО-МЕТОДИЧЕСКИЙ ЦЕНТР УПИ».
620062, РФ, Свердловская область, Екатеринбург, ул. Гагарина, 35а, оф. 2.
Тел.: 8(343)362-91-16.