Санкт-Петербургский государственный политехнический университет

Институт металлургии, машиностроения и транспорта

Кафедра конструторско-технологических инноваций

М.И. Седлер, М.Х. Седлер

СТАТИСТИЧЕСКИЕ МЕТОДЫ
В УПРАВЛЕНИИ КАЧЕСТВОМ

Учебное пособие

Санкт-Петербург
2013

УДК 519.22, 519.23, 519.25

ББК 22.172

Седлер М.И., Седлер М.Х. Статистические методы в управлении качеством: Учеб. пособие / Санкт-Петербургский государственный политехнический университет, 2013 –156 с.

Пособие соответствует ФГОС ВПО дисциплинам «Статистические методы в управлении качеством» направления подготовки бакалавров 221400.62 - Управление качеством, «Статистические методы контроля и управления качеством бизнес-процессов» направления подготовки магистров 221400.68 - Управление качеством, а также ГОС ВПО дисциплине «Статистические методы в управлении качеством» направления подготовки дипломированных специалистов 220501.65 - Управление качеством.

Рассмотрена методика статистического управления процессами с использованием семи «простых» инструментов качества. Главное внимание уделено обработке выборочных данных в программах MS Office Excel и Statistica 6.0.

Предназначено для студентов Института металлургии, машиностроения и транспорта, изучающих дисциплины «Статистические методы в управлении качеством», «Статистические методы контроля и управления качеством бизнес-процессов».

Табл. 35. Рис. 124. Библиогр.: 23 назв.

Оглавление

ВВЕДЕНИЕ

Признаки качества изделий и процессов - их свойства, обеспечивающие пригодность к выполнению определенных требований потребителя. Значения признака - возможные значения или виды проявления признака. Признак качества (случайная переменная) в каждом конкретном случае принимает значения, зависящие от случайных обстоятельств. Пример случайной величины - измеренное значение признака, являющееся результатом производственного процесса. Эти значения никогда не могут быть одинаковыми: они обладают изменчивостью. Изменчивость значений признака качества вызывается причинами изменчивости процесса, например, отклонениями в работе станка, несоответствием материала, инструмента, ошибками персонала, несоблюдением технологических параметров (температура, влажность и т.д.). Любой измеренный параметр может быть объектом статистического анализа. При выборе объекта анализа следует искать параметры, оказывающие наибольшее воздействие на качество продукции, обладающие значительной изменчивостью.

Статистическое управление процессами (SPC) является способом применения статистических методов контроля для достижения следующих целей: увеличения знаний о процессе, регулирования процесса для достижения желаемого поведения, уменьшения отклонений параметров готовой продукции, снижения трудоемкости контрольных операций путем проведения выборочного контроля. В основе методики SPC лежит применение статистических методов с использованием семи «простых» инструментов качества.

Гистограмма - столбиковый график. Высота столбика показывает число данных, попавших в каждый из интервалов. Дает наглядное изображение того, с какой частотой повторяется то или иное значение или группа значений.

Контрольная карта - график изменения параметров процесса во времени. Используется для обеспечения статистического контроля его стабильности.

Контрольный листок - таблица сбора данных и их первичной обработки (упорядочения). Используется для облегчения дальнейшего использования собранной информации.

Диаграмма Парето - разновидность столбчатой диаграммы. Позволяет распределить усилия для разрешения возникающих проблем и выявить основные причины, с которых нужно начинать действовать.

Диаграмма Ишикавы - инструмент, позволяющий выявить наиболее существенные факторы (причины), влияющие на конечный результат (следствие).

Диаграмма разброса – точечный график, позволяющий определить вид и тесноту связи между парами соответствующих переменных.

Стратификация - инструмент, позволяющий произвести отбор данных, отражающих необходимую информацию о процессе, путем расслоения (группировки) данных в зависимости от условий их получения и обработки каждой группы (страты) в отдельности.

Простые методы контроля качества рассмотрены в международном стандарте ИСО 9004-4:1993.Семь простых методов контроля качества обладают такими качествами, как простота, наглядность, визуализация результатов, они доступны для понимания персоналом любого уровня и рассчитаны на широкое применение.

Долгое время статистическая обработка информации была трудоемкой и сложной процедурой. С развитием информационных технологий эта процедура упростилась. Для решения статистических задач применяется как стандартное программное обеспечения общего назначения - табличный редактор **MS Office Excel**, так и специализированные программы. В данном учебном пособии рассмотрена обработка выборочных данных с использованием **MS Office Excel** и **Statistica 6.0.**

ОСНОВНЫЕ ПОНЯТИЯ

Партия продукции - это группа продукции, обладающая сходными характеристиками для всех единиц продукции, наработанная в один промежуток времени, из одного сырья и (обычно) наработанная на одном и том же оборудовании.

Если мы говорим о производственных процессах, то обычно статистический контроль применяют для расчетной обработки параметров (свойств) продукции (заготовок, деталей, упаковок и т.п.). С помощью этих расчетов мы можем оценить:

- стабильность производственного процесса во время наработки одной партии продукции;
- стабильность общего большого производственного процесса (разные партии продукции, полуфабрикатов и сырья);
- обнаружить случаи, когда контролируемый процесс выходит из состояния стабильности.

На рабочих чертежах проставляют размеры.

Размер — это числовое значение линейной величины (диаметра, длины, высоты и т. п.). Размеры подразделяются на номинальные, действительные и предельные.

Номинальным размером называется основной размер детали, рассчитанный с учетом ее назначения и требуемой точности. В производстве номинальные размеры не могут быть выдержаны: действительные размеры всегда в большую или меньшую сторону отличаются от номинальных размеров. Поэтому помимо номинальных размеров различают также действительные и предельные размеры деталей.

Действительный размер — размер, полученный в результате измерения готовой детали с допустимой степенью погрешности.

Допустимую неточность изготовления деталей и требуемый характер их соединения устанавливают посредством предельных размеров.

Предельными размерами называются два граничных значения, между которыми должен находиться действительный размер. Большее из этих значений называется наибольшим предельным размером, меньшее — наименьшим предельным размером. Таким образом, для обеспечения взаимозаменяемости на чертежах необходимо кроме номинального размера указывать предельные отклонения.

Предельное отклонение — это алгебраическая разность между предельными и номинальными размерами. Различают верхнее и нижнее предельные отклонения. Верхнее отклонение — это алгебраическая разность между наибольшим предельным размером и номинальным размером. Нижнее отклонение — алгебраическая разность между наименьшим предельным размером и номинальным размером.

Действительное отклонение — алгебраическая разность между действительным и номинальным размерами. Деталь считают годной, если действительное отклонение проверяемого размера находится между верхним и нижним отклонениями.

Допуск - разность между наибольшим и наименьшим предельными размерами или абсолютная величина алгебраической разности между верхним и нижним отклонениями.

Допустим, что для детали на чертеже задан размер $19h7_{-0,021}$. Это означает, что действительный размер должен колебаться только в пределах 18,979 – 19,000 мм.

Разброс значений (вариация) - различие значений какого-либо признака у разных единиц совокупности за один и тот же промежуток времени.

Случайная вариация и Систематическая вариация. Несмотря на то, что практически невозможно произвести две одинаковые детали в одной партии, их свойства будут очень близки, поскольку они были произведены из одной партии сырья, на одном станке (линии) в одну и ту же смену, одним и тем же рабочим.

Появление небольших изменений в размерах деталей, которое происходит без изменений во внешних факторах (изменение сырья, изменения в станке, человеческое вмешательство и т.п.) называется случайной вариацией. Ее контролировать невозможно.

Изменение в размерах деталей, вызванное внешним фактором (сбой в станке, скажем) называется систематической вариацией.

ГЕНЕРАЦИЯ СЛУЧАЙНЫХ ЧИСЕЛ В MS OFFICE EXCEL

С развитием компьютерной техники, даже самые сложные статистические расчеты оперативно выполняются современными программами. Предположим, что в результате контроля размера $19h7_{-0,021}$ выборки 50 значений получен ряд случайных размеров.

В **MS Office Excel** можно сгенерировать выбор у нормального распределения, имитирующую 50 значений размеров деталей. Для этого нужно открыть новую книгу и загрузить надстройку **Пакет анализа.**

Для этого:

1. Нажмите кнопку **“Office” - Параметры Excel** (рис. 1)

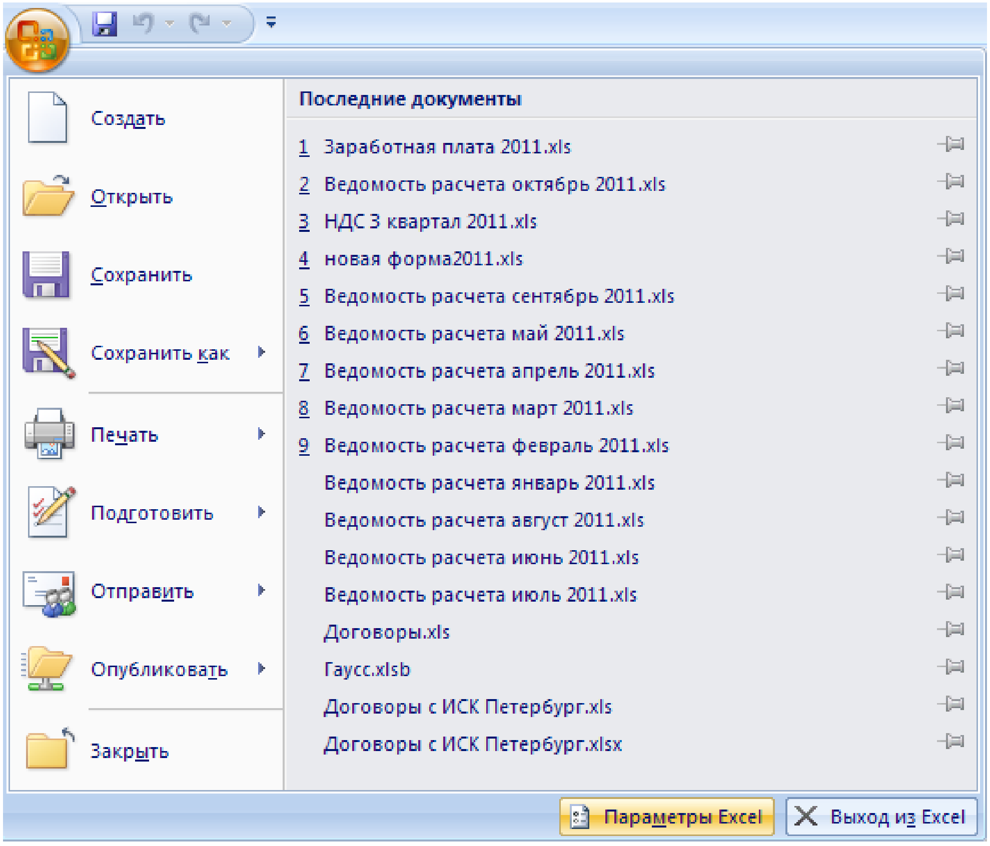


Рис. 1

2. Перейдите на вкладку **Надстройки** (рис. 2)

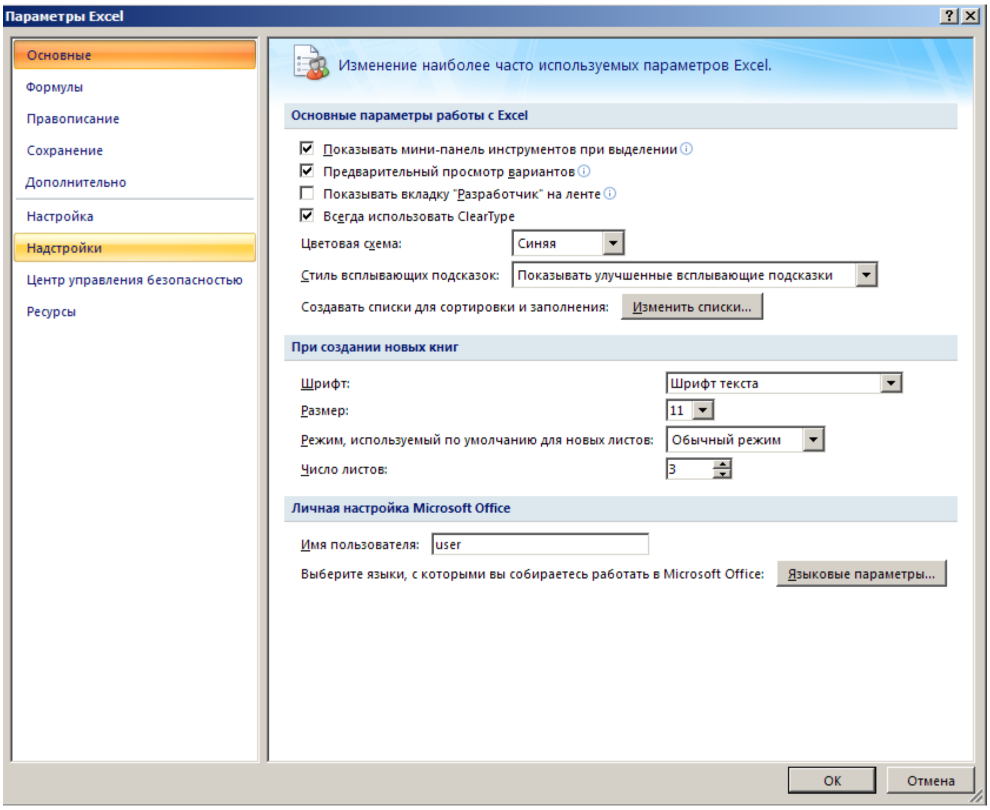


Рис. 2

3. Выберите **Пакет анализа**, нажмите **Перейти** (рис. 3)

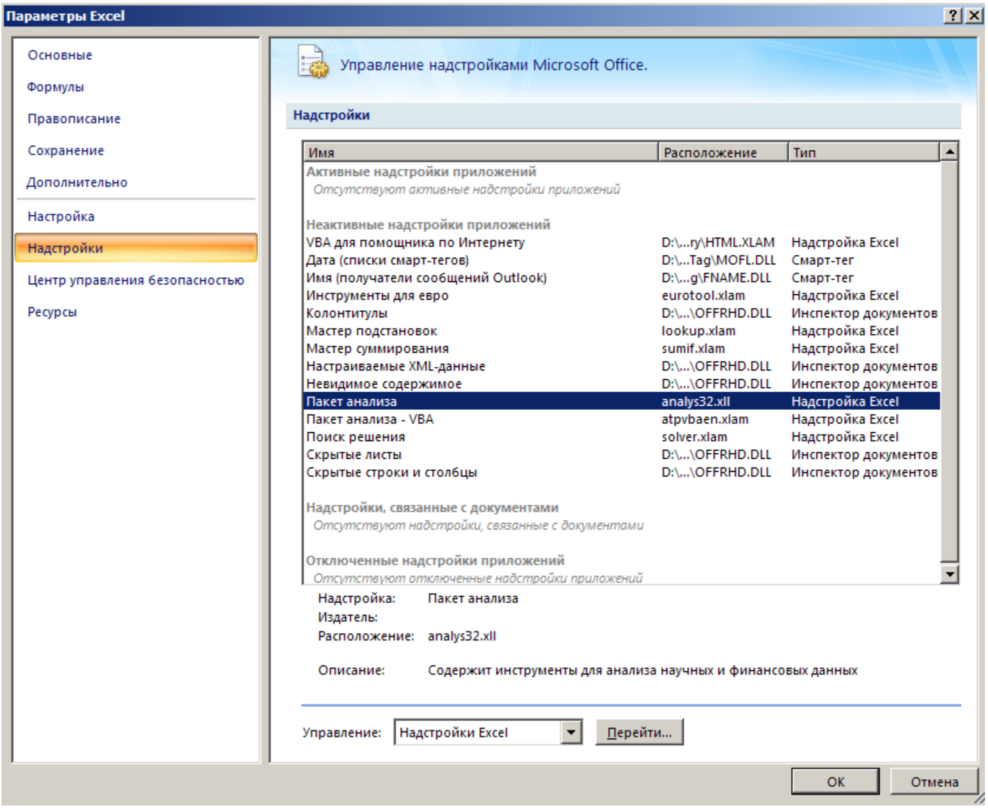


Рис. 3

4. Выберите в доступных надстройках **Пакет анализа** и нажмите **ОК** (рис. 4).

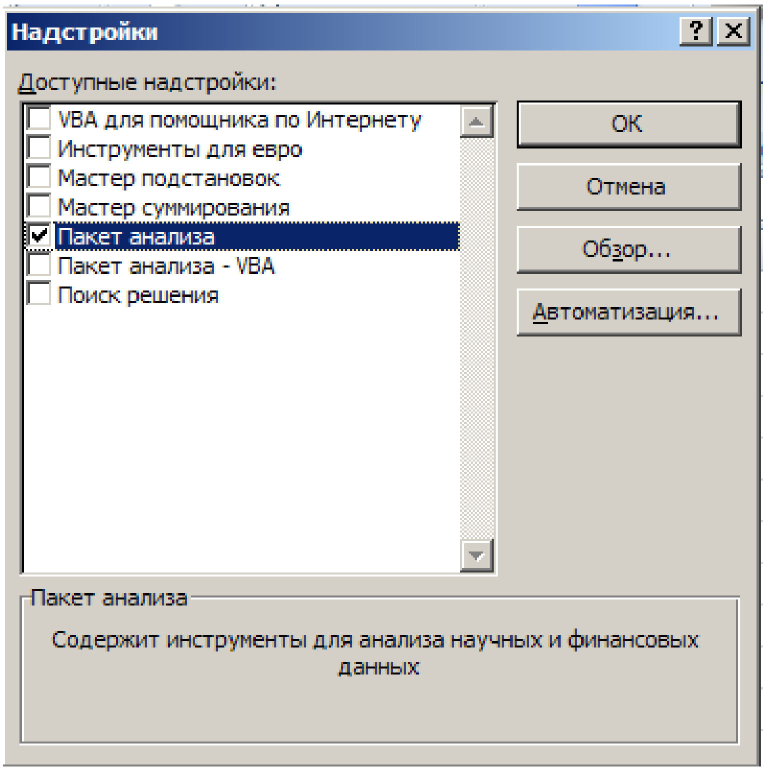


Рис. 4

В случае если компонент не установлен, программа предложит установить компонент. Для этого потребуется дистрибутив программы MS Office Excel. После загрузки надстройки **Пакет анализа** на вкладке **Данные** становится доступной команда **Анализ данных.**

Генерация выборки нормального распределения осуществляется в следующей последовательности.

1. **Шаг 1.** Зайдите в меню **Сервис - Анализ данных - Генерация случайных чисел.** Нажмите **ОК** (рис. 5).

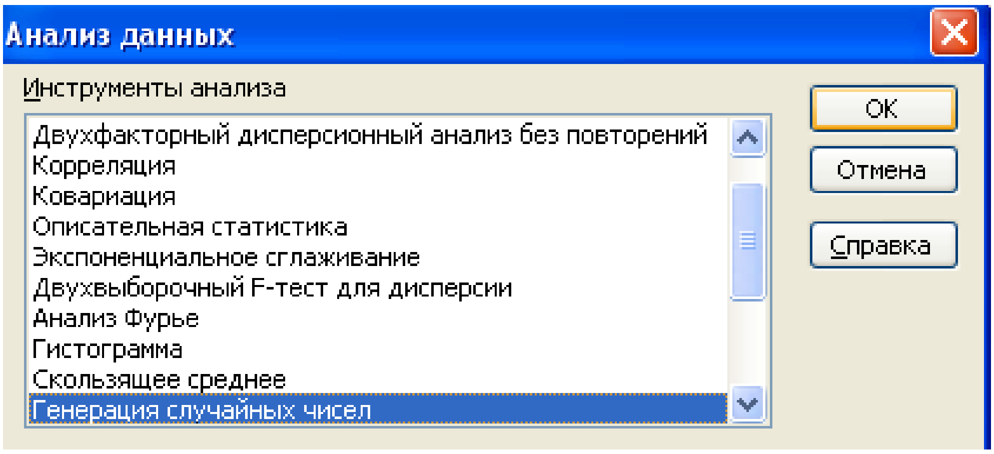


Рис. 5

2. **Шаг 2.** В появившемся окне выберите число переменных 1, число случайных чисел – 50, распределение выберите из раскрывающегося списка «Нормальное» с параметрами Среднее значение = 18,990, Стандартное отклонение = 0,003 и сгенерируйте выборку в выходном интервале A1. Нажмите **ОК** (рис. 6).

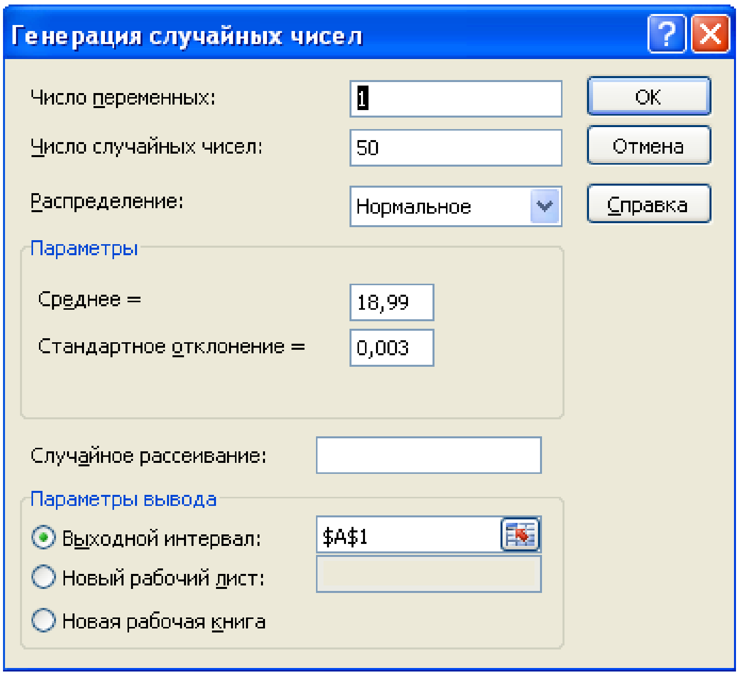


Рис. 6

В результате произведенных действий будет сгенерирован столбец данных– результаты измерений длины каждой детали.

<u>Задание.</u> В MS Office Excel с помощью надстройки Анализ данных сформируйте выборку нормального распределения 50 индивидуальных значений - результатов контроля размера деталей с параметрами, указанными в варианте задания (Приложение 1).

ОБРАБОТКА ВЫБОРОЧНЫХ ДАННЫХ В MS OFFICE EXCEL

Начальный этап первичной обработки данных связан с поиском максимального $X_{\max}$ и минимального $X_{\min}$ значений выборки $X = X_1, X_2, \ldots, X_n$, а также размаха варьирования:

$$R = X_{\max} - X_{\min} \tag{1}$$

Для определения размаха необходимо расположить единицы выборки в возрастающем, или убывающем порядке. Для этого выберите в меню пиктограмму **Сортировка** и задайте сортировку значений по возрастанию (рис. 7).

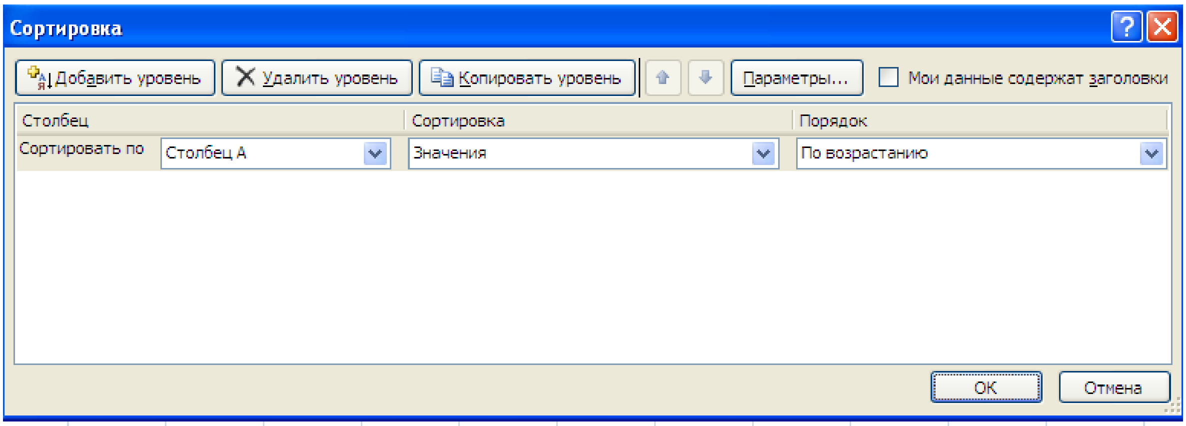


Рис. 7

Следующий этап первичной обработки заключается в группировке данных и их графическом представлении.

При группировке промежуток $X_{\min}, X_{\max}$ разбивают на m интервалов и подсчитывают число выборочных значений n_j, где j = 1,2,...,m, которые попали в j-й интервал. Как правило, число интервалов m = 6...20.

Для определения оптимального числа групп можно использовать формулу Стерджесса:

$$m = 1 + 3{,}322 \times \lg n \tag{2}$$

где n - общее число единиц выборки, равное 50.

$m = 6{,}6$. Примем число интервалов равным 7.

Величина интервала группировки $\Delta_j = b_j - a_j$ рассчитывается по формуле

$$\Delta_j = \frac{R}{m}, \tag{3}$$

где R – размах варьирования.

Ширина интервалов для всего ряда должна быть одинаковой; $\Delta_j = const$.

За середину первого интервала $\Delta_1 = b_1 - a_1$ может быть принято значение x_1, равное $X_{\min}$; за середину последнего интервала значение x_m, равное $X_{\max}$. Тогда размах варьирования определяется по формуле:

$$R = x_m - x_1 = 0{,}017 \tag{4}$$

$\Delta_j = 0{,}002$

Каждое отдельное значение n_j должно быть однозначно отнесено к определенному интервалу. Каждое значение n_j, попадающее на границу интервалов $\Delta_j = b_j - a_j$ и $\Delta_{j+1} = b_{j+1} - a_{j+1}$, что бывает сравнительно редко, должно быть причислено к интервалу по принятому правилу.

Такие значения рекомендуют причислять к интервалам одним из трех способов:

1. Принять равным 1 и причислить к нижнему интервалу $\Delta_j = b_j - a_j$;
2. Принять равным 1 и причислить к верхнему интервалу $\Delta_{j+1} = b_{j+1} - a_{j+1}$;
3. Принять равным 1/2 и причислить значение 1/2 к нижнему интервалу $\Delta_j = a_j, b_j$, значение 1/2 к верхнему интервалу $\Delta_{j+1} = b_{j+1} - a_{j+1}$.

Создайте таблицу расчета границ интервалов и произведите расчет частоты попадания значений сгенерированной выборки в интервалы (таблица 1).

Таблица 1

Интервал	**Начало**	**Середина**	**Конец**	**Частота попаданий**
	18,980	18,981	18,983	1
2	18,983	18,984	18,986	4
3	18,986	18,987	18,989	11
4	18,989	18,990	1,992	21
5	18,992	18,993	18,995	9
6	18,995	18,996	18,998	4
7	18,998	18,999	19,001	0

ГИСТОГРАММЫ

Наиболее наглядной формой графического представления группировки является гистограмма. Гистограмма используется:

- для показа характера изменчивости;
- сообщения визуальной информации о ходе процесса;
- принятия решения о фокусе усилий по улучшению.

Наиболее распространенные формы гистограмм приведены на рисунке 8.

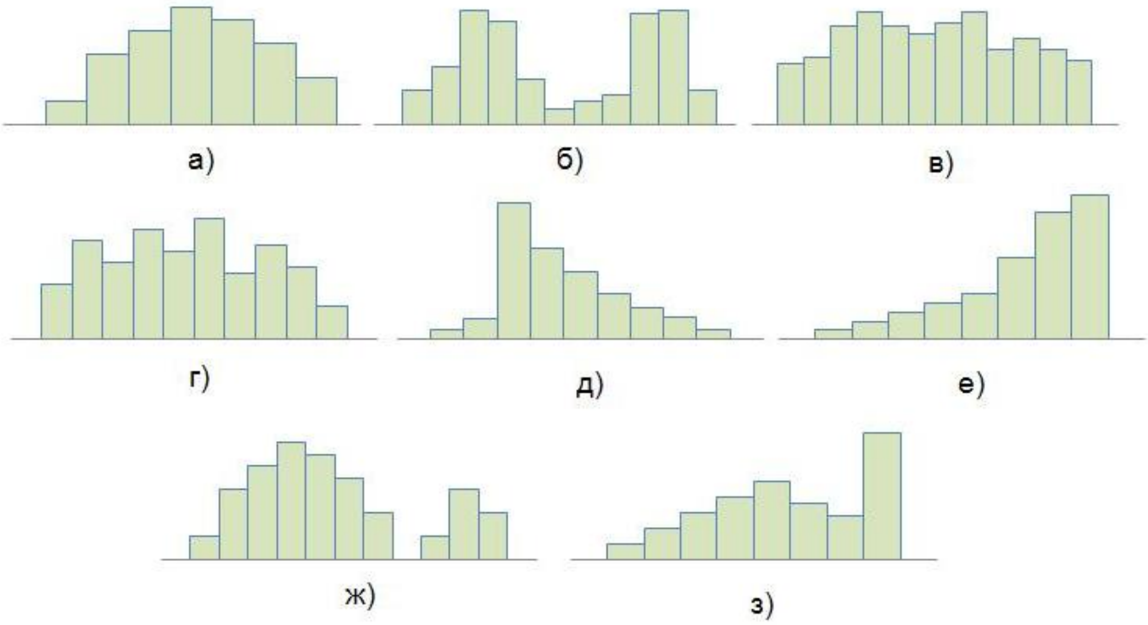


Рис. 8

Колоколообразное распределение (а) – симметричная форма с максимумом примерно в середине интервала изменения изучаемого параметра; характерно для распределения параметра по нормальному закону, при равномерном влиянии на него различных факторов. Отклонения от колоколообразной формы могут указывать на наличие доминирующих факторов или нарушений методики сбора данных (например, включение в выборку данных, полученных в других условиях).

Распределение с двумя пиками, или двухвершинное, (б) характерно для выборки, объединяющей результаты двух процессов или условий работы. Например, когда анализируются результаты измерений размеров деталей после обработки, такая гистограмма будет иметь место, если в одну выборку объединены измерения деталей при разных настройках инструмента или при использовании разных инструментов либо станков. При анализе полученных данных используют различные схемы стратификации с целью выделения различных процессов или условий.

Распределение типа плато (в) имеет место для тех же условий, что и предыдущая гистограмма. Особенностью данной выборки является то, что в ней объединено несколько распределений, в которых средние значения незначительно отличаются между собой. Целесообразно провести анализ последовательно выполняемых операций, применить альтернативные процедуры реализации операций для уменьшения вариабельности условий процессов и их результатов. При анализе полученных данных полезно также применение метода стратификации (расслоения) данных.

Распределение гребенчатого типа (г) – регулярно чередующиеся высокие и низкие значения. Этот тип обычно указывает на ошибки измерений, на ошибки в способе группировки данных при построении гистограммы или на систематическую погрешность в способе округления данных. Менее вероятна альтернатива того, что это один из вариантов распределения типа плато. Необходимо проанализировать процедуры сбора данных и построения гистограммы, прежде чем рассматривать возможные характеристики процесса, которые могли бы вызывать такую структуру.

Скошенное распределение (д) имеет асимметричную форму с пиком, расположенным не в центре данных, и с «хвостами» распределения, которые резко спадают с одной стороны и мягко – с другой. Иллюстрация на рисунке называется положительно скошенным распределением, потому что длинный «хвост» простирается вправо к уменьшающимся значениям. Отрицательно скошенное распределение имело бы длинный «хвост», простирающийся влево к уменьшающимся значениям. Такая форма гистограммы указывает на отличие распределения изучаемого параметра от нормального распределения. Оно может быть вызвано преобладающим влиянием какого-либо фактора на разброс значений параметра, например, при механической обработке это может быть влияние точности заготовок или оснастки на точность обработанных деталей.
Такие распределения возможны, так как обусловлены природой получения выборок.

Усеченное распределение (е) это часто гладкие, колоколообразные распределения, у которых посредством некоторой внешней силы (отбраковка, 100%-ный контроль или перепроверка) часть распределения изъята или усечена. Усеченное имеет асимметричную форму, при которой пик находится на краю или вблизи от края данных, а распределение с одной стороны обрывается очень резко и имеет плавный «хвост» с другой стороны.

Распределение с изолированным пиком (ж) имеет небольшую, отдельную группу данных в дополнение к основному распределению. Как и распределение с двумя пиками, эта структура представляет собой некоторую комбинацию и предполагает, что работают два различных процесса. Однако маленький размер второго пика указывает на ненормальность, на что-то такое, что не происходит часто или регулярно. При анализе такого распределения необходимо обратить внимание на условия, сопутствующие данным в маленьком пике: нельзя ли обособить конкретное время, оборудование, источник входных материалов, процедуру, оператора и т. д. Такие маленькие изолированные пики в сочетании с усеченным распределением могут быть следствием отсутствия достаточной эффективности отбраковки дефектных изделий. Возможно, что маленький пик представляет ошибки в измерениях или переписывании данных.

Распределение с пиком на краю (з) имеет большой пик, присоединенный к гладкому в остальном распределению. Такая форма существует тогда, когда протяженный «хвост» гладкого распределения был обрезан и собран в одну-единственную категорию на краю диапазона данных. Кроме того, это указывает на неаккуратную запись данных.

Для этого в таблице (рис. 8) выделите столбец частот, перейдите на вкладку **Вставка,** выделите пиктограмму **Гистограмма** и **выберите вид гистограммы** (рис. 9).

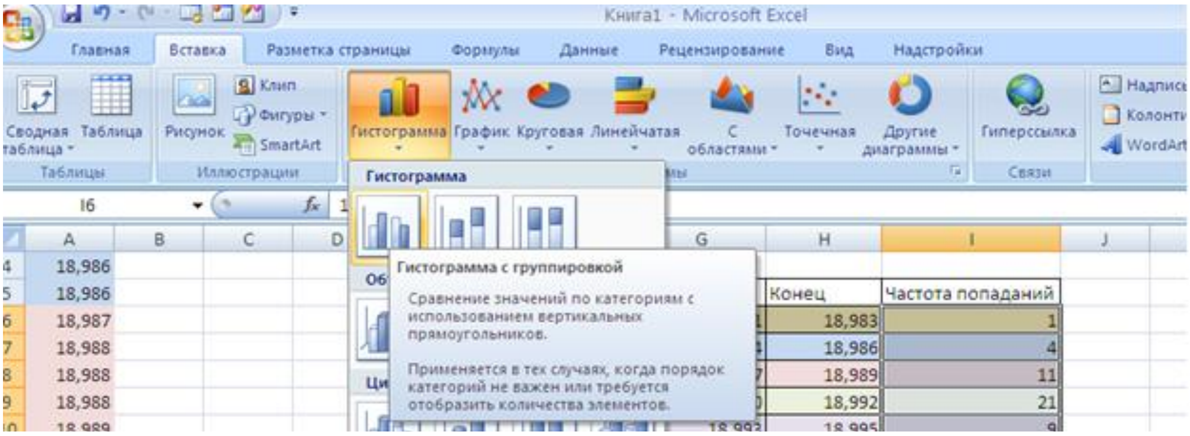


Рис. 9

В результате произведенных действий будет построена гистограмма распределения частот (рис. 10).

Рис. 10

<u>Задание</u>
В MS Office Excel создайте таблицу группировки 50 индивидуальных значений - результатов контроля размера деталей. По результатам расчетов постройте гистограмму распределения частот.

ФОРМЫ ГРАФИЧЕСКОГО ПРЕДСТАВЛЕНИЯ СГРУППИРОВАННЫХ ДАННЫХ

Наиболее наглядной формой графического представления группировки является *гистограмма распределения относительных частот*, которая представляет собой график решетчатой функции

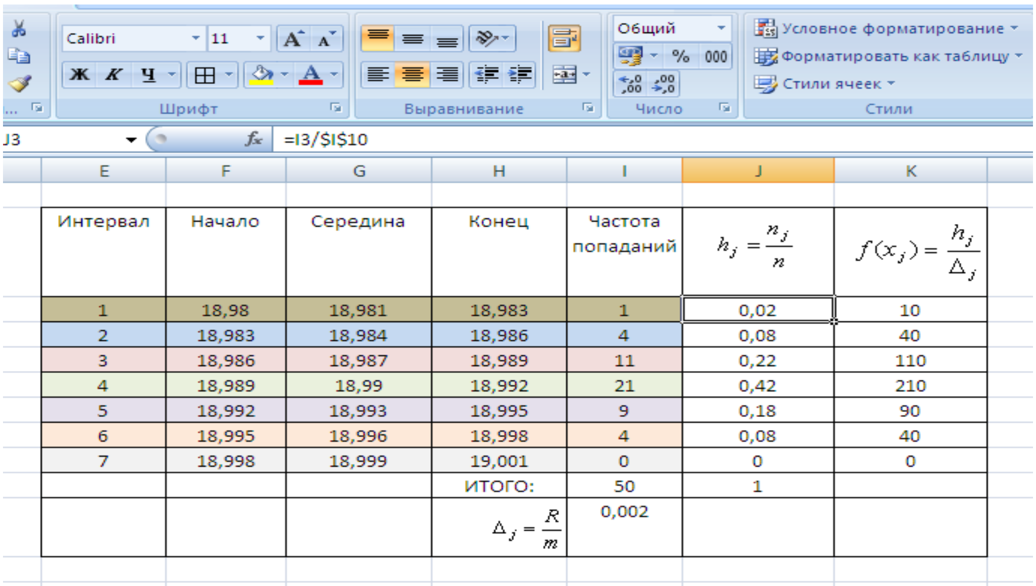
$$f(x_j) = \frac{h_j}{\Delta_j},\ x_j \in \Delta_j,\ j = 1,2,\ldots,m, \qquad (5)$$

где h_j - относительная частота попаданий вычисляется по формуле:

$$h_j = \frac{n_j}{n}, \qquad (6)$$

где Δ_j - интервал группировки.

В таблицу сгруппированных данных MS Excel добавьте столбцы, для которых по формулам рассчитайте значения данных столбцов h_j, $f(x_j)$ (рис. 11), постройте гистограмму (рис. 12).



Интервал	Начало	Середина	Конец	Частота попаданий	$h_j = \frac{n_j}{n}$	$f(x_j) = \frac{h_j}{\Delta_j}$
1	18,98	18,981	18,983	1	0,02	10
2	18,983	18,984	18,986	4	0,08	40
3	18,986	18,987	18,989	11	0,22	110
4	18,989	18,99	18,992	21	0,42	210
5	18,992	18,993	18,995	9	0,18	90
6	18,995	18,996	18,998	4	0,08	40
7	18,998	18,999	19,001	0	0	0
			ИТОГО:	50	1	
			$\Delta_j = \frac{R}{m}$	0,002		

Рис. 11

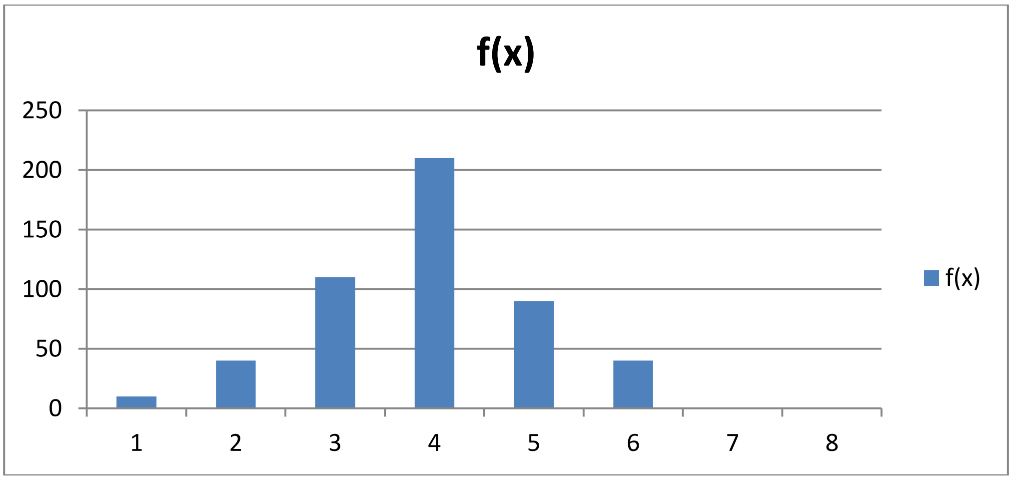


Рис. 12

Другая форма графического представления группированных данных - полигон частот, который изображается в виде графика - ломаной линии, с абсциссами $x_j, j = 1,2,...,m$, определяемыми серединами интервалов группировки $\Delta_j = b_j - a_j$ и ординатами, равными частотам попадания наблюдений в интервалы группировки - n_j, или относительным частотам

$$h_j = \frac{n_j}{n},\% \qquad (7)$$

В таблице сгруппированных данных MS Excel рассчитайте значения данных столбца $h_j = \frac{n_j}{n},\%$ (рис. 13). Для построения полигона частот выделите данные столбца относительных частот, перейдите на вкладку **Вставка,** выберите пиктограмму **График** и **задайте вид графика** (рис. 14).

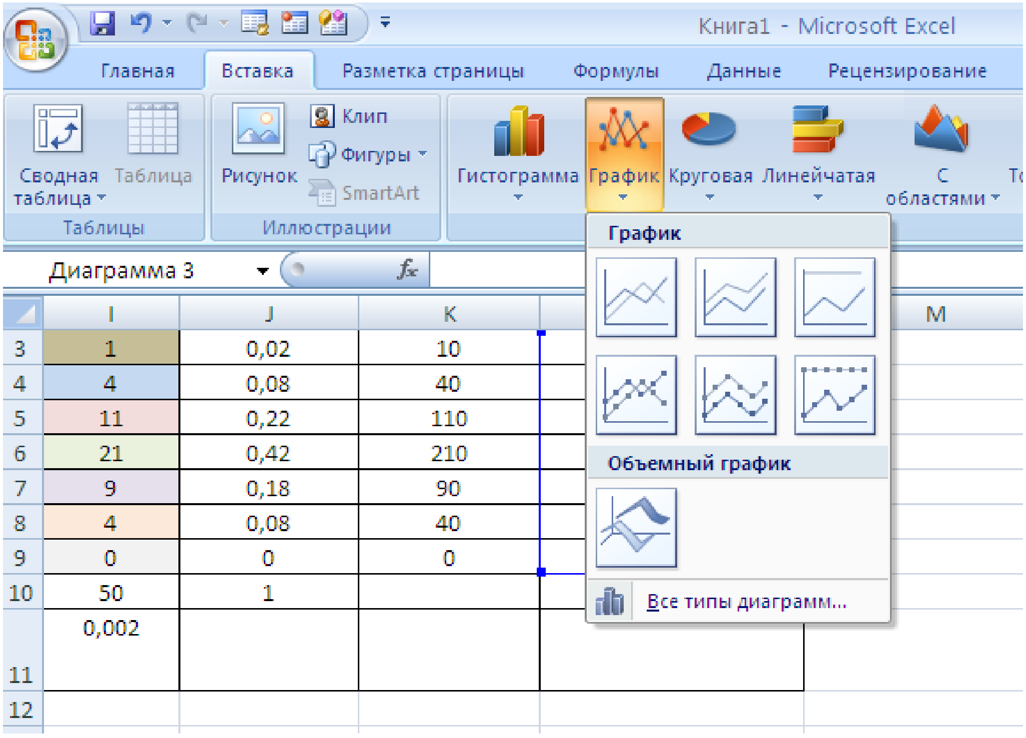


Рис. 13

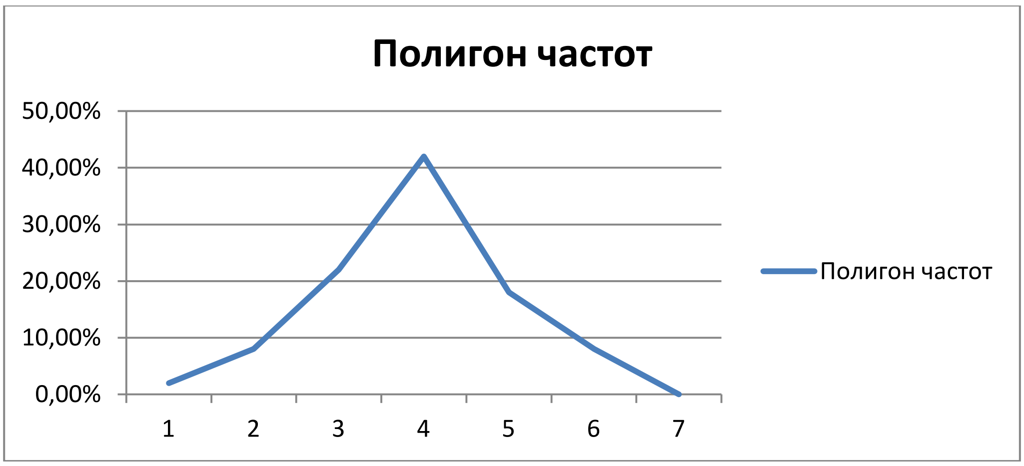


Рис. 14

Еще одна форма представления данных *Полигон накопленных частот (кумулятивная кривая)* - график ломаной линии, с абсциссами $x_j,\ j=1,2,\ldots,m$, определяемыми серединами интервалов группировки $\Delta_j = b_j - a_j$, и ординатами, равными сумме накопленных частот $\sum_{j=1}^{m} n_j$, или сумме относительных накопленных частот

$$\sum_{J=1}^{m} h_j = \frac{\sum_{j=1}^{m} n_j}{n}, \% \qquad (8)$$

В таблице сгруппированных данных MS Excel рассчитайте значения данных столбца $\sum_{J=1}^{m} h_j = \frac{\sum_{j=1}^{m} n_j}{n}$,%. По полученным данным постройте полигон накопленных частот (рис. 15).

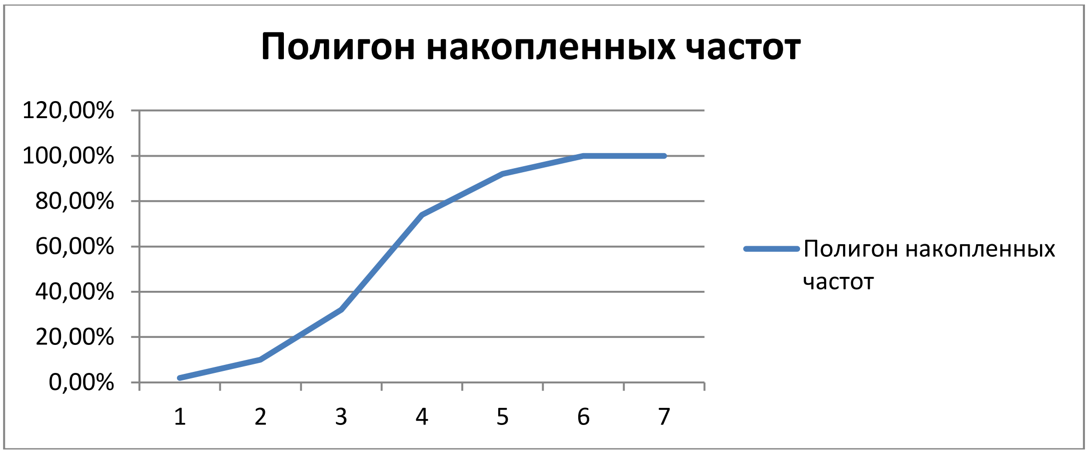


Рис. 15

<u>Задание</u>
В таблицу сгруппированных данных MS Office Excel добавьте столбцы, рассчитайте значения и постройте гистограмму распределения относительных частот, полигон частот и полигон накопленных частот.

Среднее значение выборки случайной величины, или выборочное среднее вычисляется по формуле средней арифметической

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{9}$$

Для расчета среднего значения выберите свободную ячейку и введите в нее статистическую функцию: =СРЗНАЧ(A1:A50). В результате вычислений получено значение выборочного среднего $\bar{x}$ = 18,990.

Выборочное среднее для упорядоченного ряда сгруппированных данных вычисляется по формуле средней арифметической взвешенной:

$$\bar{x} = \frac{\sum_{j=1}^{m} (x_j \times f_j)}{\sum_{j=1}^{m} f_j} \tag{10}$$

по таблице сгруппированных данных.

Для расчета выборочного среднего сгруппированных данных выберите свободную ячейку и введите в нее следующую формулу: =СУММПРОИЗВ(G3:G9;K3:K9)/(СУММ(К3:К9).

В результате вычисления получено значение выборочного среднего $\bar{x}$ =18,9897.

Дисперсия - статистическая характеристика ряда наблюдаемых значений, показывающая, как тесно группируются отдельные значения вокруг средней арифметической или как они рассеиваются вокруг этой средней.

Алгебраическая сумма отклонений отдельных значений x_i от средней арифметической $\bar{x}$ равна нулю. Поэтому она не пригодна в качестве меры рассеяния. Поэтому за меру рассеяния принимают сумму квадратов отклонений отдельных значений от средней арифметической, деленную на количество наблюдений. Эту меру называют дисперсией и обозначают через s^2

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{11}$$

Для расчета дисперсии выберите свободную ячейку и введите в нее статистическую функцию: =ДИСП(A1:A50).

В результате вычислений получено значение дисперсии s^2 =0,000013.

Для вычисления дисперсии упорядоченного ряда сгруппированных данных пользуются следующей формулой

$$s^2 = \frac{\Delta}{n} \sum_{j=1}^{m} \frac{(x_j - \bar{x})^2}{\Delta_j} n_j \tag{12}$$

по таблице сгруппированных данных.

Поскольку при группировке выборка была разбита на одинаковые интервалы, формула вычисления дисперсии упорядоченного ряда преобразуется к виду

$$s^2 = \frac{1}{n} \sum_{j=1}^{m} (x_j - \bar{x})^2 n_j \tag{13}$$

Для расчета дисперсии упорядоченного ряда сгруппированных данных добавьте в таблицу сгруппированных данных столбец значений

$$\sum_{j=1}^{m}(x_j - \bar{x})^2 n_j$$

В ячейки столбца введите формулу: =(G4-M11)^2*I4.

Для расчета дисперсии введите в свободную ячейкуформулу:=N10/I10.

В результате вычислений получено значение дисперсии упорядоченного ряда сгруппированных данных s^2 =0,000011.

Среднее квадратическое отклонение s часто применяют вместо дисперсии s^2. Оно имеет ту же размерность, что и средняя арифметическая $\bar{x}$.

$$s = \sqrt{s^2} \tag{14}$$

Для расчета среднего квадратического отклонения выберите свободную ячейку и введите в нее **статистическую функцию**: =СТАНДОТКЛОН(A1:A50). В результате вычислений получено значение $s = 0{,}0036$

Для расчета среднего квадратического отклонения упорядоченного ряда сгруппированных данных выберите свободную ячейку и введите в нее **формулу: =КОРЕНЬ(M12).** В результате вычислений получено значение $s = 0{,}0033$.

<u>Задание</u>

В MS Office Excel рассчитайте статистические характеристики для 50 индивидуальных значений - результатов контроля размера деталей и для упорядоченного ряда сгруппированных данных.

ОБРАБОТКА ВЫБОРОЧНЫХ ДАННЫХ В STATISTICA 6.0

Для обработки данных выборки откройте файл MS Excel «Построение гистограмм и расчет параметров», сохраните его в формате Excel 97-2003 и откройте сохраненный файл в программе Statistica 6.0 (рис. 16).

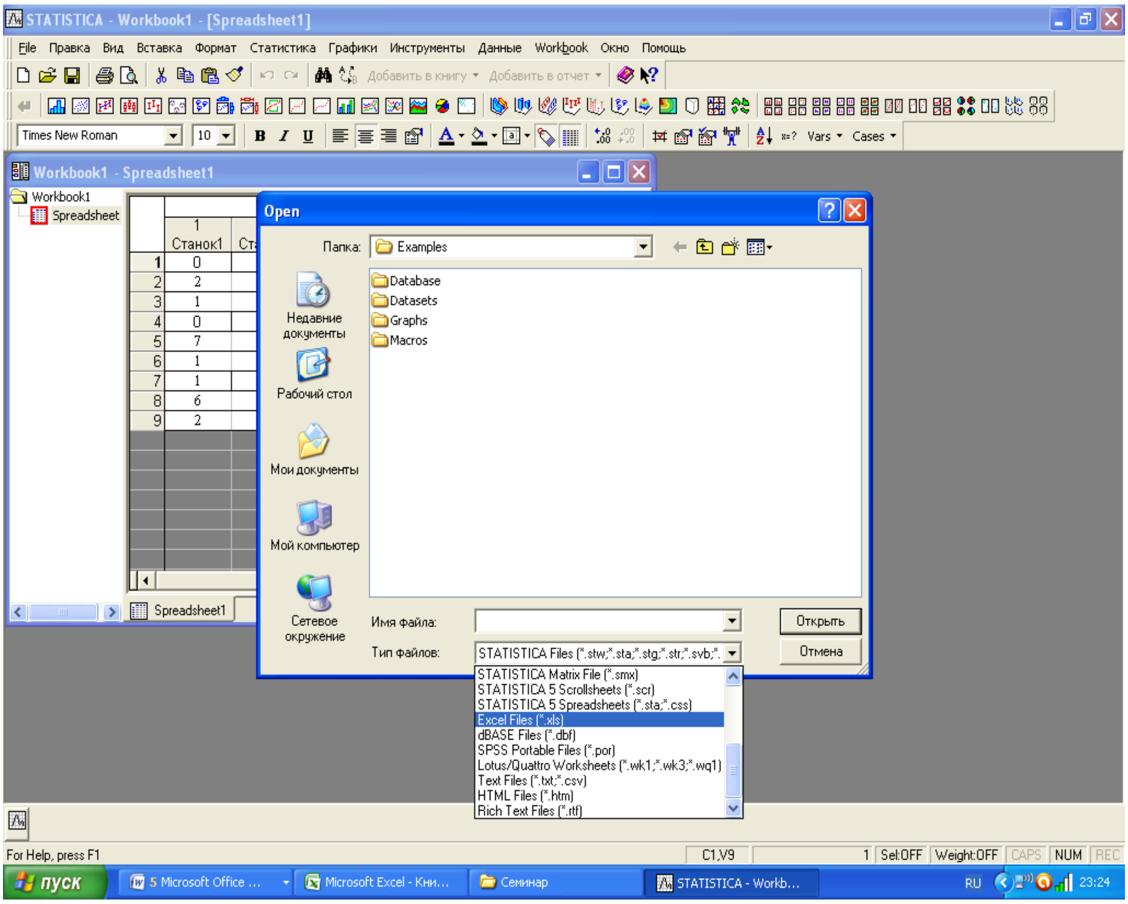


Рис. 16

Для этого настройте схему импорта данных. Нажмите **OK** (рис. 17).

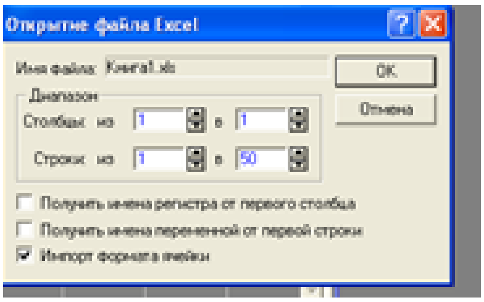


Рис. 17

Выделите в меню пиктограмму **Статистика,** в раскрывшемся списке **Основная статистика/таблицы** (рис. 18).

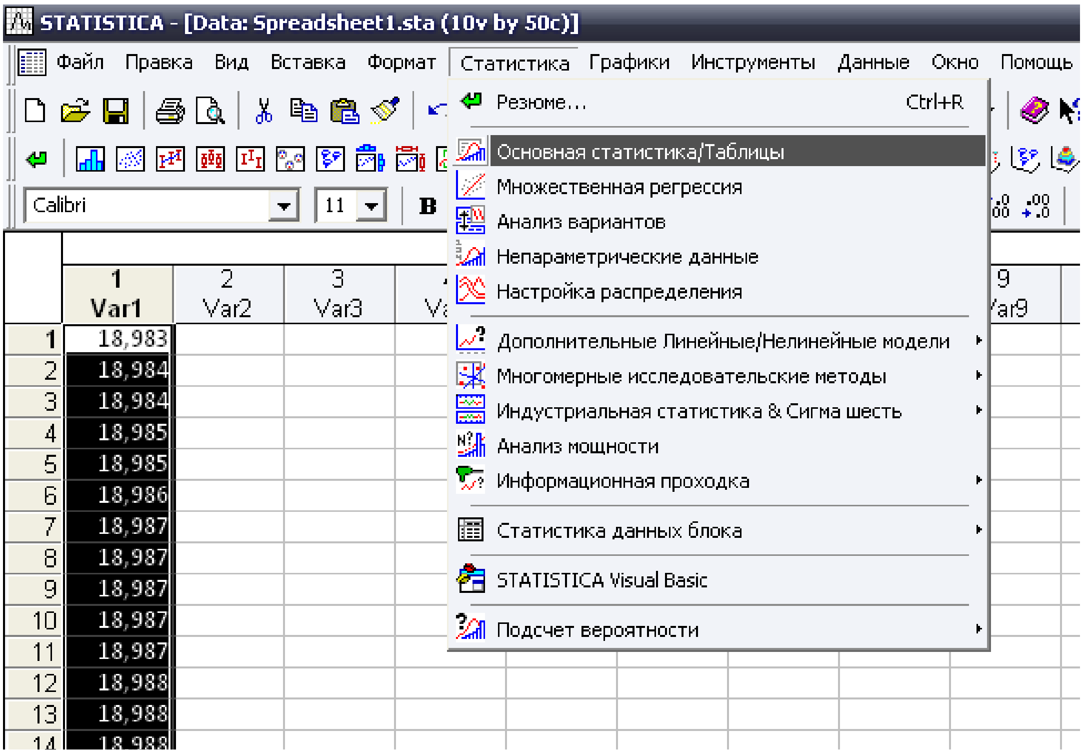


Рис. 18

Далее в раскрывшемся списке выделите **Descriptive statistics.** Нажмите **OK** (рис.19).

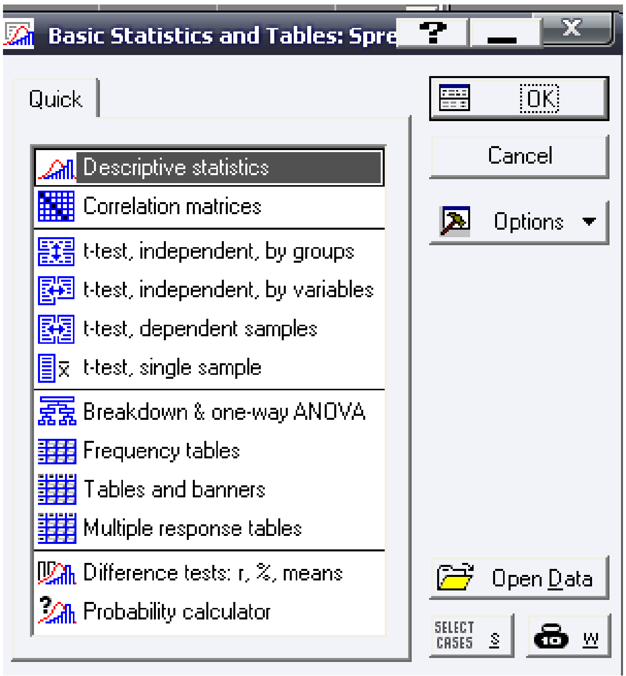


Рис. 19

В открывшемся окне задайте переменную для анализа Var1. Нажмите **ОК** (рис. 20).

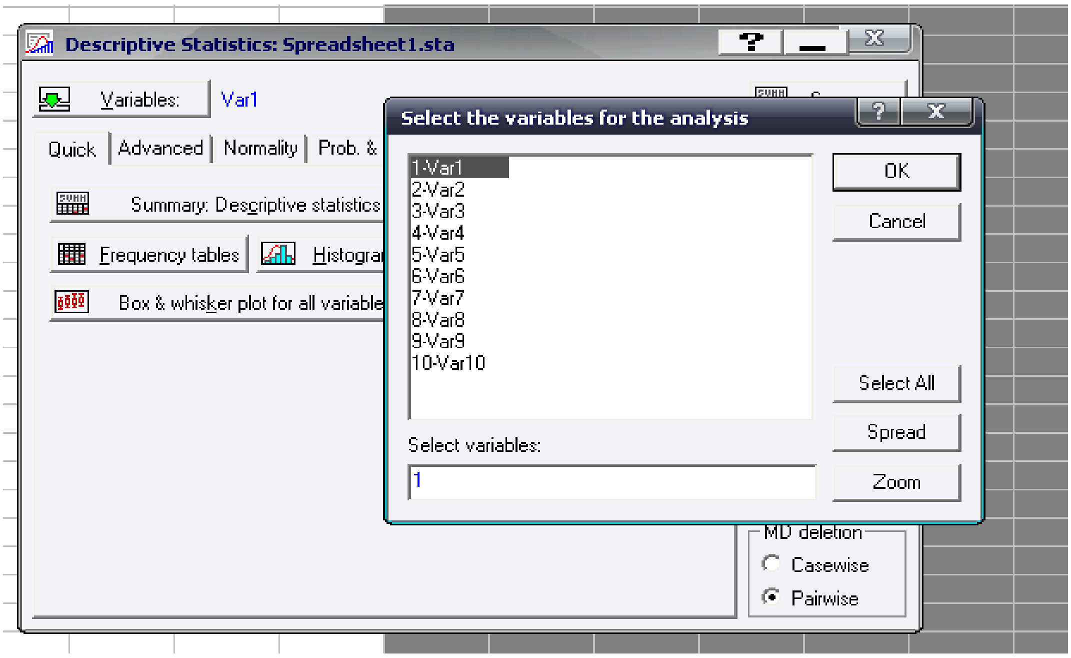


Рис. 20

Перейдите на вкладку **Advanced** и установите доверительную вероятность 95% (рис. 21).

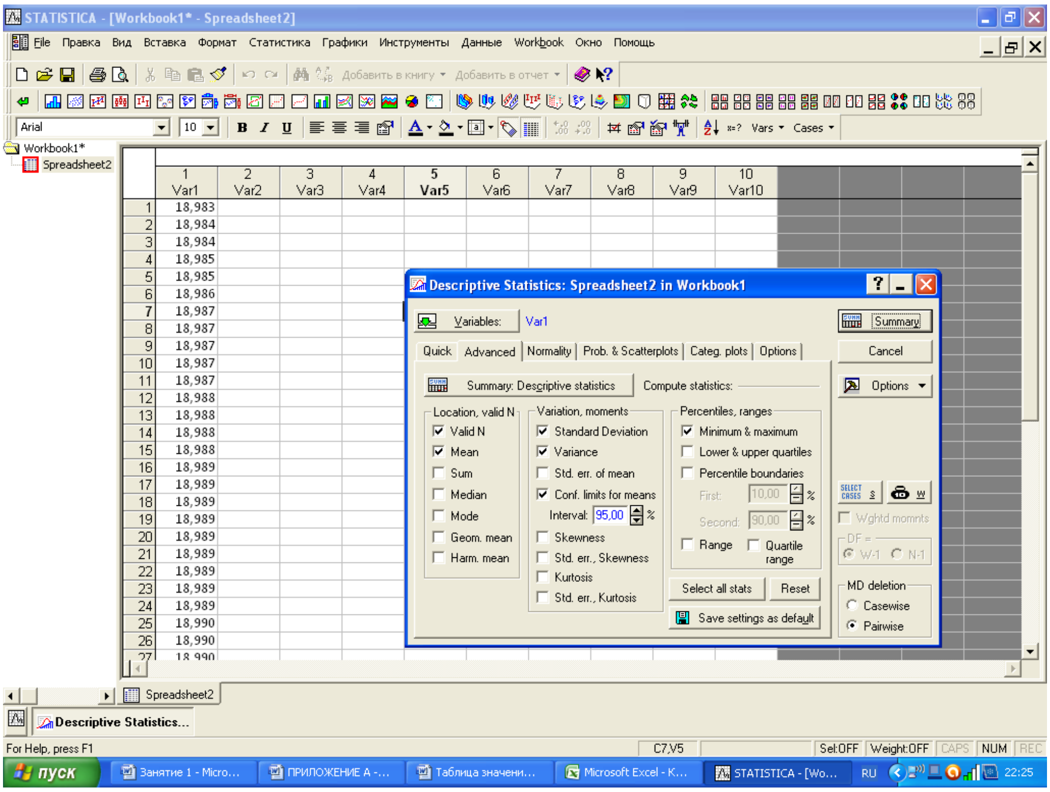


Рис. 21

Перейдите на вкладку **Normality** и задайте число интервалов, равное 7 (рис. 22).

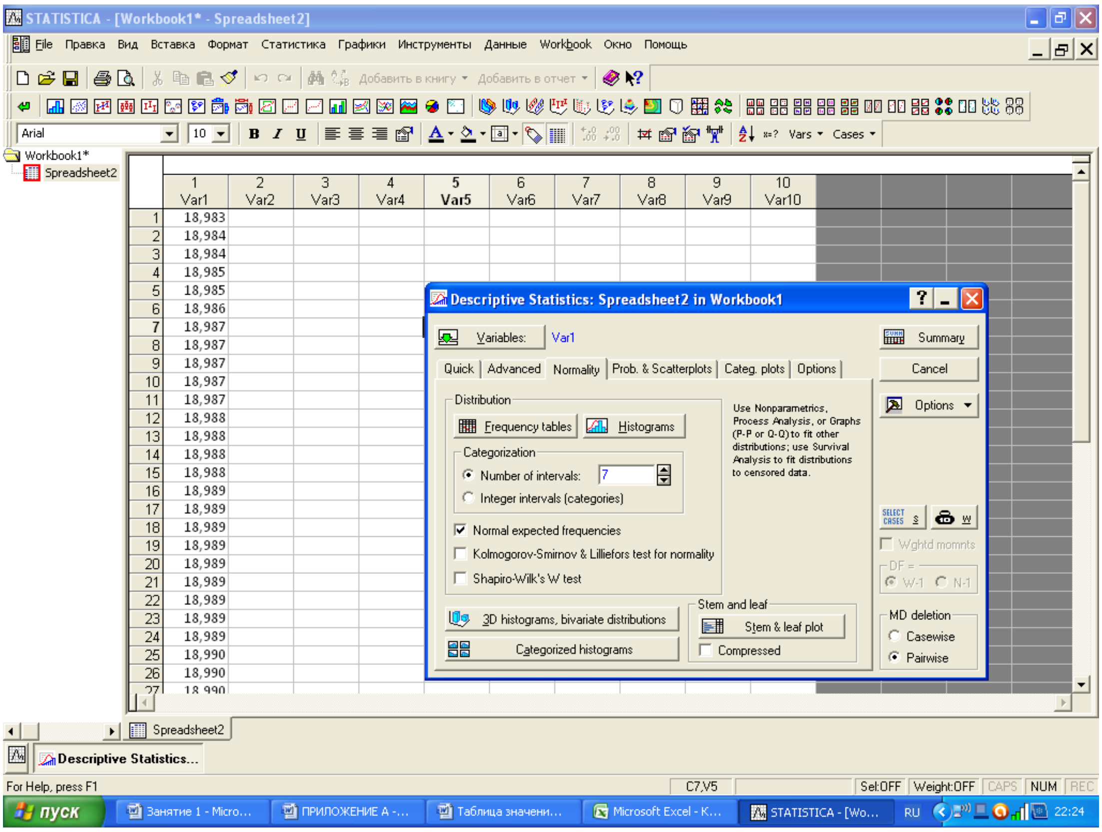


Рис. 22

Нажмите на кнопку **Summary.** Сравните полученные результаты с результатами, вычисленными в MS Excel (рис. 23).

Variable	Descriptive Statistics (Spreadsheet2 in Workbook1)							
	Valid N	Mean	Confidence -95,000%	Confidence +95,000%	Minimum	Maximum	Variance	Std.Dev.
Var1	50	18,99014	18,98912	18,99115	18,98336	18,99867	0,000013	0,003572

Рис. 23

При нажатии на кнопку **Frequency tables** будет сформирована таблица сгруппированных данных (рис. 24).

Category	Frequencytable: Var1 (Spreadsheet2 in Workbook1)									
	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All	Expected Count	Cumulative Expected	Percent Expected	Cumulative % Expected
18,98000<x<=18,98250	0	0	0,00000	0,0000	0,00000	0,0000	0,81248	0,81248	1,62497	1,62497
18,98250<x<=18,98500	4	4	8,00000	8,0000	8,00000	8,0000	2,94592	3,75841	5,89184	7,51681
18,98500<x<=18,98750	7	11	14,00000	22,0000	14,00000	22,0000	7,74723	11,50563	15,49445	23,01126
18,98750<x<=18,99000	16	27	32,00000	54,0000	32,00000	54,0000	12,72446	24,23009	25,44892	48,46018
18,99000<x<=18,99250	14	41	28,00000	82,0000	28,00000	82,0000	13,05880	37,28889	26,11760	74,57778
18,99250<x<=18,99500	3	44	6,00000	88,0000	6,00000	88,0000	8,37430	45,66319	16,74860	91,32638
18,99500<x<=18,99750	3	47	6,00000	94,0000	6,00000	94,0000	3,35423	49,01742	6,70846	98,03485
18,99750<x<=19,00000	3	50	6,00000	100,0000	6,00000	100,0000	0,83846	49,85588	1,67691	99,71176
Missing	0	50	0,00000		0,00000	100,0000				

Рис. 24

При нажатии на кнопку **Histograms** будет построена гистограмма (рис. 25).

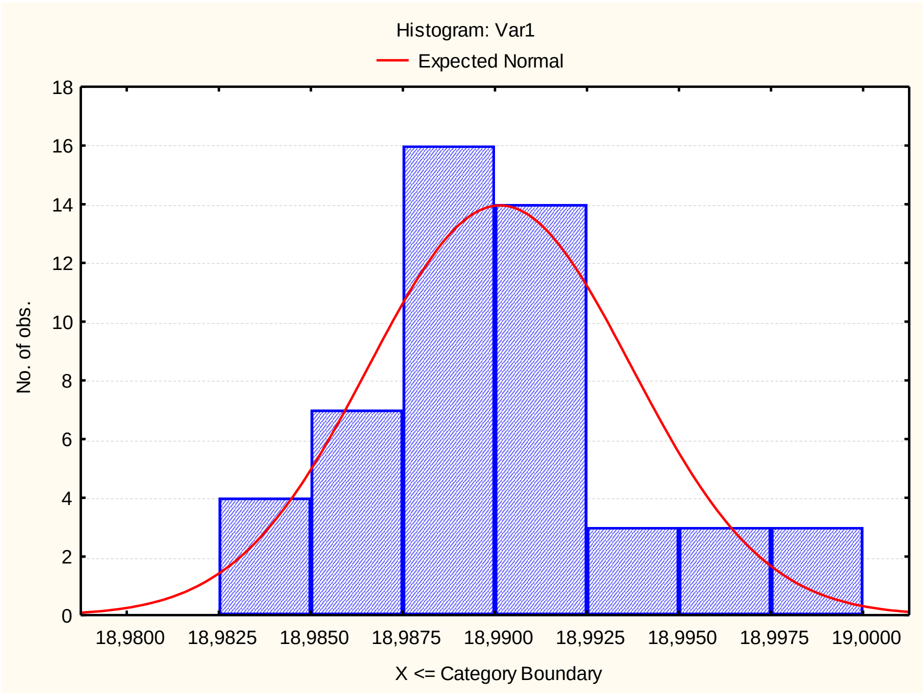


Рис. 25

Полученные результаты можно сохранить в файлы, или добавить в отчет. Для добавления данных в отчет нажмите на кнопку **Options,** выберите в раскрывшемся списке **Вывод**…(рис. 26).

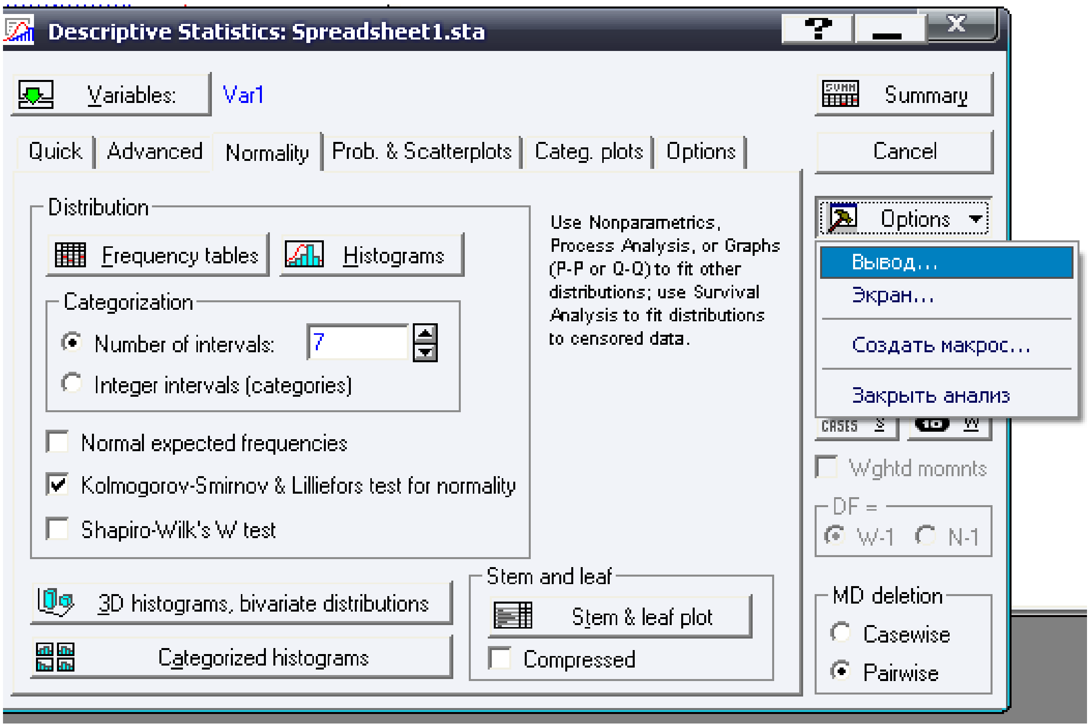


Рис. 26

В открывшемся окне задайте необходимость вывода в отчет и выберите тип отчета. Нажмите **ОК** (рис. 27).

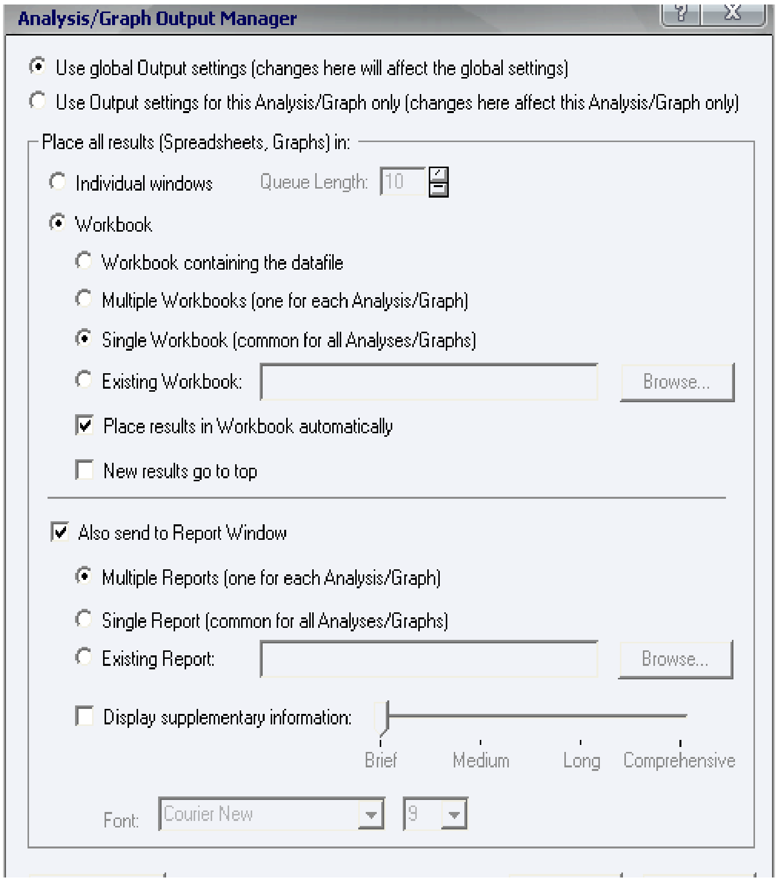


Рис. 27

<u>Задание</u>
Осуществите импорт выборочных данных из файла MS Office Excel в программу Statistica 6.0. Произведите анализ 50 индивидуальных значений - результатов контроля размера деталей. Результаты анализа добавьте в отчет.

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Одним из непрерывных распределений, имеющим основополагающую роль в математической статистике, является нормальное, или гауссово, распределение. Значительное число законов распределения, встречающихся в практике контроля качества промышленной продукции, аппроксимируются кривой нормального распределения.

Непрерывная случайная величина *X*называется распределенной по **нормальному закону**, если ее плотность распределения описывается функцией

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \tag{15}$$

где $m = const,\ \sigma = const > 0$

Свойства нормального распределения:

– Нормальная кривая имеет колоколообразную форму, симметричную относительно точки $x = m$, с точками перегиба, абсциссы которых отстоят от m на $\pm\sigma$.

– Область определения этой функции: (-∞, +∞).

– ***f***(***x***) > 0 при любом ***x.***

– при ***x = m*** $f(x) = \frac{1}{\sigma\sqrt{2\pi}}$ - точка максимума.

– ***f***(***x – m)*** = ***f***(***m – x***), то есть график симметричен относительно прямой ***x = m.***

– Точки перегиба:

$x_1 = m + \sigma$

$x_2 = m - \sigma$

– Нормальное распределение определяется параметрами: математическим ожиданием:

$$M(X) = m \tag{16}$$

– дисперсией

$$D(X) = \sigma^2 \tag{17}$$

– стандартным отклонением:

$$\sigma(X) = \sqrt{D(X) =}\sigma \tag{18}$$

График плотности вероятности нормального распределения, называемый **нормальной кривой,** показывает, что для нормально распределенной случайной величины вероятность отклонения от среднего значения быстро уменьшается с ростом величины отклонения.

Нормированное нормальное распределение

Формула плотности нормально распределенной непрерывной случайной величины описывает целое семейство кривых, зависящих от двух параметров и , которые могут принимать любые значения, поэтому возможно бесконечно много нормально распределенных совокупностей.

Чтобы избежать неудобств, связанных с расчетами для каждого конкретного случая по достаточно сложной формуле, используют так называемое нормированное (или стандартное) нормальное распределение **с параметрами** $m = 0$, $\sigma = 1$.

Это распределение получается, если **пронормировать** нормально распределенную величину X по формуле:

$$u = \frac{x-m}{\sigma} \qquad (19)$$

Плотность распределения вероятностей нормированного нормального распределения записывается в виде: $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}, -\infty < u < +\infty$

На кривой нормированного нормального распределения (рис. 28) указаны в процентах доли площадей соответствующих отмеченным значениям нормированного отклонения и, по отношению к общей площади под кривой, равной 1 (100%). Эти площади определяют вероятности попадания случайной величины в соответствующие интервалы.

Рис. 28

Значения ординат нормальной кривой приведены в специальных таблицах значений функции Гаусса. Значения $\varphi(u)$ для некоторых характерных нормированных отклонений представлены в таблице 2.

Таблица 2

Нормированное отклонение,	0	±0,5	±1,0	±2,0	±3,0
Ордината нормальной кривой,	0,399	0,352	0,242	0,054	0,004

Вероятность попадания в заданный интервал
Функция Лапласа

Найдем вероятность попадания случайной величины, распределенной по нормальному закону, в заданный интервал.

$$P(a < X < b) = \int_a^b f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

Обозначим $\frac{x-m}{\sigma\sqrt{2}} = t; \quad \frac{a-m}{\sigma\sqrt{2}} = \alpha; \quad \frac{b-m}{\sigma\sqrt{2}} = \beta;$

Тогда

$$P(a < X < b) = \frac{1}{\sigma\sqrt{2\pi}}\int_{\alpha}^{\beta} e^{-t^2}\sigma\sqrt{2}dt = \frac{1}{2}\Phi(\beta) - \frac{1}{2}\Phi(\alpha)$$

Т.к. интеграл $\int e^{-t^2}dt$ не выражается через элементарные функции, то вводится в рассмотрение функция стандартного нормального закона распределения.

Функцию распределения для произвольных параметров можно выразить через нормированную функцию Лапласа. Тогда

$$P(a < X < b) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right) \quad (20)$$

В таблице 3 приведены полученные вероятности того, что нормально распределенная случайная величина отклонится от своего среднего значения m не более чем на $\pm 0{,}5\sigma, \pm\sigma, m \pm 2\sigma, m \pm 3\sigma$.

Вероятности попадания нормально распределенной случайной величины в заданный интервал

Таблица 3

Границы интервала, $m \pm x$	$m + 0{,}5\sigma$	$m \pm \sigma$	$m \pm 2\sigma$	$m \pm 3\sigma$
Вероятность попадания в интервал	0,3829	0,6827	0,9545	0,9973

На практике можно считать, что ***все*** возможные значения нормально распределенной случайной величины лежат в интервале (*m* - 3*σ*, *m* + 3*σ*).

Правило «трех сигм»: ***если случайная величина распределена нормально, то модуль ее отклонения от x = m не превосходит 3σ.***

КРИТЕРИЙ СОГЛАСИЯ ПИРСОНА

Чрезвычайно важной задачей сравнительной статистики является сравнение эмпирического и теоретического (или гипотетического) распределений.

Если между ними имеется согласие, то можно заключить, что эмпирическое распределение вызвано теми же причинами, которые лежат в основе теоретического распределения. Пригодной для проверки статистикой является величина χ^2.

Приближенный критерий проверки гипотезы об отсутствии зависимости основывается на том факте, что малые различия между наблюдаемыми и ожидаемыми числами поддерживают эту гипотезу, а большие отклонения свидетельствуют против нее.

Нуль-гипотеза: в основе выборки лежит предполагаемое теоретическое распределение $F_0(x)$. Согласно альтернативной гипотезе выборка принадлежит к неизвестному распределению $F(t)$.

Мы рассматриваем независимую выборку $X = X_1, X_2, \ldots.., X_n$, обозначая неизвестную функцию распределения $F(t)$. Нас интересует вопрос о том, согласуются ли данные наблюдений $X_1, X_2, \ldots.., X_n$ с простой гипотезой H_0: $F(t) = F_0(t)$, где $F_0(t)$- некоторая конкретная фиксированная функция распределения.

Мерой расхождения между теоретическими и эмпирическими вероятностями принимается величина χ^2.

Если выборка разделена на m интервалов, причем наблюдаемые значения расположены в отдельных интервалах случайным образом независимо друг от друга, то значение χ^2 определяется как сумма по всем интервалам квадратов разностей между наблюдаемыми и ожидаемыми частотами, поделенных на ожидаемые частоты.

$$\chi^2 = \sum_{j=1}^{m} \frac{(h_{эмп} - h_{теор})^2}{h_{теор}} \qquad (21)$$

где $h_{эмп}$ - частота попаданий случайной величины в интервал группировки $\Delta_j = a_j, b_j$ (наблюдаемая частота);

$h_{теор}$- теоретическое значение частоты для соответствующего интервала (ожидаемая частота);

m - число интервалов группировки.

Статистика χ^2 называется статистикой хи-квадрат Пирсона. В случае, если нуль-гипотеза верна, $\chi^2 = 0$, если неверна $\chi^2 \longrightarrow \infty$.

Минимальное число значений, необходимое для вычисления статистики, называется *числом степеней свободы* данной статистики,

$$k = (m - 1 - r), \qquad (22)$$

где m - число интервалов выборки;

r - число параметров, оцениваемых по выборке (для нормального распределения r =2).

Практический смысл теоремы Пирсона состоит в том, что при *достаточно большом объеме* выборки можно рассчитать статистику χ^2, зная значения разностей эмпирических и теоретических значений частот.

То обстоятельство, что поведение χ^2 существенно различно в зависимости от того верна или нет нуль-гипотеза, дает возможность построить критерий для ее проверки.

Для проверки гипотезы вероятность ошибки (уровень значимости) задается заранее.

При этом нуль-гипотезу можно:

- отклонить, когда $p < 0{,}01$, или принять, когда $p > 0{,}10$;
- воздержаться, когда $0{,}01 < p < 0{,}10$;
- повторить эксперименты для получения большего числа данных.

Действия по принятию (или отвержению) нуль - гипотезы состоят в следующем.

1. Подстановкой имеющихся выборочных данных в формулу для вычисления значения функции χ^2, которое затем сравнивается с табличным значением для $k = (m - 1 - r)$ степеней свободы и принятого уровня значимости.

2. Если значение функции χ^2 окажется больше табличного значения h, то нуль-гипотеза *отвергается* (при этом говорят, что выборка обнаруживает *значимое отклонение* от нуль - гипотезы);

3. Если значение функции χ^2 окажется меньше или равно табличного значения h, то нуль-гипотеза *принимается* (говорят, что выборка *совместима* с нуль - гипотезой).

ПРОВЕРКА ГИПОТЕЗЫ О НОРМАЛЬНОМ РАСПРЕДЕЛЕНИИ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ ПО КРИТЕРИЮ ПИРСОНА

Эмпирическое распределение задано в виде последовательности равноотстоящих вариант и соответствующих им частот.

X	X_1	X_2	…………	X_N
N	N_1	N_2	…………..	N_N

Требуется, используя критерий Пирсона, проверить гипотезу о том, что генеральная совокупность распределена нормально.

Если выборка содержит все возможные результаты измерений, то эти результаты представляют собой **генеральную совокупность.** Отметим, что генеральная совокупность измерений может содержать как бесконечное число элементов (как в данном примере), так и конечное число элементов. Обычно выборка содержит малую часть генеральной совокупности и поэтому лишь приближенно характеризует свойства генеральной совокупности.

Для того, чтобы при заданном уровне значимости проверить гипотезу о нормальном распределении генеральной совокупности, надо:

1. Вычислить непосредственно выборочное среднее $\bar{x}$ и выборочное среднее квадратичное отклонение s .

2. Вычислить теоретические частоты:

$$n_i' = \frac{n\Delta_i}{s}\varphi(u_i), \tag{23}$$

где n - объем выборки (сумма всех частот),

Δ_i - интервал группировки,

s - среднее квадратичное отклонение;

$$u_i = \frac{x_i - \bar{x}}{s} \tag{24}$$

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \tag{25}$$

Пояснение происхождения формулы (23).

Плотность общего нормального распределения

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2} /(2\sigma^2) \tag{26}$$

При $m = 0$ и $\sigma = 1$ получим плотность нормированного распределения:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2} / 2,$$

или изменив обозначение аргумента

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2} / 2.$$

Положив $u = (x - m)/\sigma$ имеем:

$$\varphi(u)=\frac{1}{\sqrt{2\pi}}e^{-(x-m)^2}/(2\sigma^2) \quad (27)$$

Сравнивая (26) и (27), можно заключить

$\varphi(x)=\frac{1}{\sigma}\varphi(u)$.

Если математическое ожидание m и среднее квадратичное отклонение σ неизвестны, то в качестве оценок этих параметров принимают соответственно выборочную среднюю $\bar{x}_в$ и выборочное среднее квадратичное отклонение $\sigma_в$. Тогда:

$$\varphi(x)=\frac{1}{\sigma}\varphi(u), \quad (28)$$

где $u=(x-m)/\sigma$.

Пусть x_i - середина i-го интервала длиной h. Тогда вероятность попадания X в этот интервал приближенно равна произведению длины интервала на значение плотности распределения $f(x)$ в любой точке интервала и, в частности при $x=x_i$

$$P_i=hf(x_i)=h\frac{1}{\sigma_в}\varphi(u_i) \quad (29)$$

Следовательно, выравнивающая частота

$$n_i'=nP_i=\frac{nh}{\sigma_в}\varphi(u_i) \quad (30)$$

где $u_i=(x_i-\bar{x}_в)/\sigma_в$, т.е. получена формула 23.

Проверьте гипотезу о нормальном распределении рассмотренной ранее выборки 50 значений результатов контроля размера $19h7_{-0,021}$:

1. Для середин интервалов сгруппированных данных рассчитайте аргументы функции Гаусса.

$$u_i=(x_i-\bar{x})/\sigma$$

2. Для каждого аргумента по таблице значений функции Гаусса определите $\varphi(u_i)$. Таблица значений функции Гаусса приведена в Приложении 1.

3. Для каждого значения функции $\varphi(u_i)$ рассчитайте теоретические частоты по формуле (23):

$$n_i'=\frac{n\Delta_i}{s}\varphi(u_i)$$

4. Сравните эмпирические и теоретические частоты с помощью критерия Пирсона. Для этого рассчитайте слагаемые и вычислите значение функции по формуле 31 (таблица 4):

$$\chi^2{}_{набл} = \sum_{j=1}^{m} \frac{(n_j - n_j^{/})^2}{n_j^{/}} \qquad (31)$$

Таблица 4

				$n_i' = \frac{n\Delta_i}{s}\varphi(u_i)$	$\chi^2{}_{набл} = \sum_{j=1}^{m} \frac{(n_j - n_j^{/})^2}{n_j^{/}}$
18,981	1	-2,61	0,0132	0,4	1,07
18,984	4	-1,71	0,0925	2,7	0,62
18,987	11	-0,81	0,2874	8,6	0,65
18,99	21	0,09	0,3973	12,0	6,67
18,993	9	0,99	0,2444	7,3	0,38
18,996	4	1,89	0,0669	2,0	2,15
18,999	0	2,79	0,0081	0,2	0,23
$\chi^2{}_{набл}$					11,77

1. По таблице критических точек распределения χ^2 (Приложения 2) для заданного уровня значимости α и числа степеней свободы $k = (m-3)$ (m - число групп выборки) определите критическую точку. $\chi^2{}_{кр}(\alpha; k)$.

Если $\chi^2{}_{набл} < \chi^2{}_{кр}$ - нет оснований опровергать гипотезу о нормальном распределении генеральной совокупности. Другими словами, эмпирические и теоретические частоты различаются незначимо. В противном случае гипотезу отвергают. Для k=4:

$\chi^2{}_{кр}(0{,}01;4) = 13{,}3$

Таким образом, с вероятностью менее 5% можно утверждать, что выборочное распределение соответствует нормальному распределению.

Добавьте к гистограмме распределения эмпирических частот график нормального распределения теоретических частот (рис. 29).

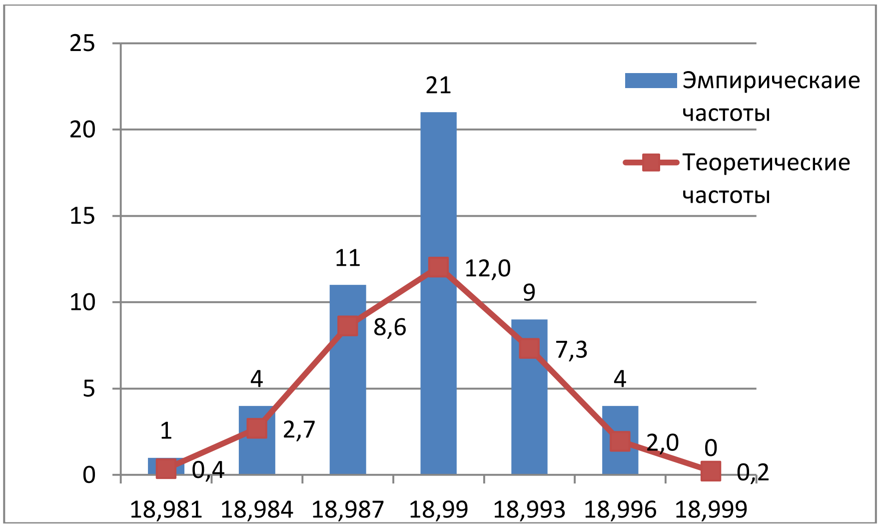


Рис. 29

Полученные данные можно проверить на соответствие нормальному закону с помощью статистической функции проверки по критерию Пирсона. Для этого в свободную ячейку введите функцию =ХИ2ТЕСТ(H4:H10;P4:P10).

В результате проверки получено значение вероятности на соответствие нормальному распределению, равное p = 6,7%.

Это связано с тем, что утверждения теоремы Пирсона относятся к пределам при $n \longrightarrow \infty$. *На практике* мы имеем дело лишь с выборками ограниченного объема. Поэтому, применяя вышеописанный критерий, необходимо проявлять осторожность. Согласно рекомендациям, применение критерия дает хорошие результаты, когда все ожидаемые частоты $np_i^0 \geq 10$. **Если же какие-то из этих чисел малы, рекомендуется, укрупняя некоторые группы, перегруппировать данные таким образом, чтобы ожидаемые частоты всех групп были не меньше десяти.** Если число m достаточно велико (имеет порядок нескольких десятков), то порог для ожидаемых частот может быть понижен до 5 или даже до 3.

<u>Задание</u>

В MS Office Excel проверьте гипотезу о нормальном распределении генеральной совокупности по критерию Пирсона для 50 индивидуальных значений - результатов контроля размера деталей. Добавьте к гистограмме распределения эмпирических частот график нормального распределения теоретических частот.

ГЕНЕРАЦИЯ ВЫБОРКИ НОРМАЛЬНОГОРАСПРЕДЕЛЕНИЯ В STATISTICA 6.0

С помощью генератора случайных чисел сформируйте выборку нормального распределения. Для этого откройте новый файл базы данных. Задайте число переменных 4, количество значений 30 (рис. 30). Параметры распределения: среднее значение $m = 15$, стандартное отклонение $\sigma = 3$.

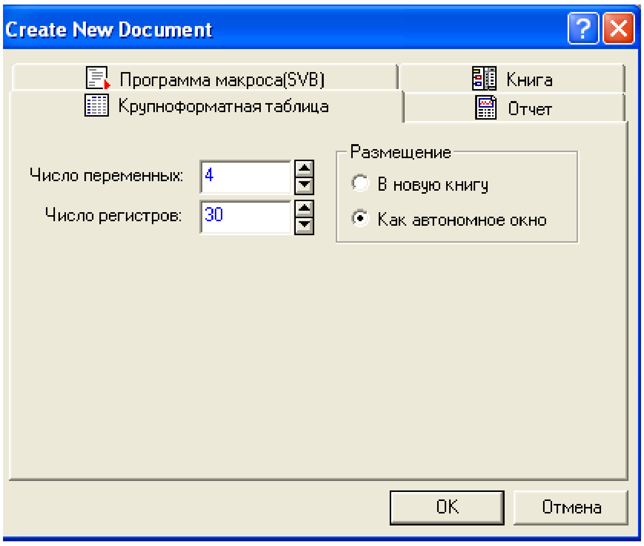


Рис. 30

Щелкните два раза левой клавишей мыши по заголовку первого столбца. Сгенерируйте в первом столбце выборку нормального распределения по формуле (32):

$$Z = \frac{x - m}{\sigma} \qquad (32)$$

Для этого в открывшемся окне нажмите на кнопку Function, в открывшемся окне Function Browser, найдите в списке функцию RndNormal, щелкните по ней два раза левой клавишей мыши, в области задания функций в скобках укажите ссылку на первый столбец и в открывшемся диалоговом окне нажмите **ОК** (рис. 31).

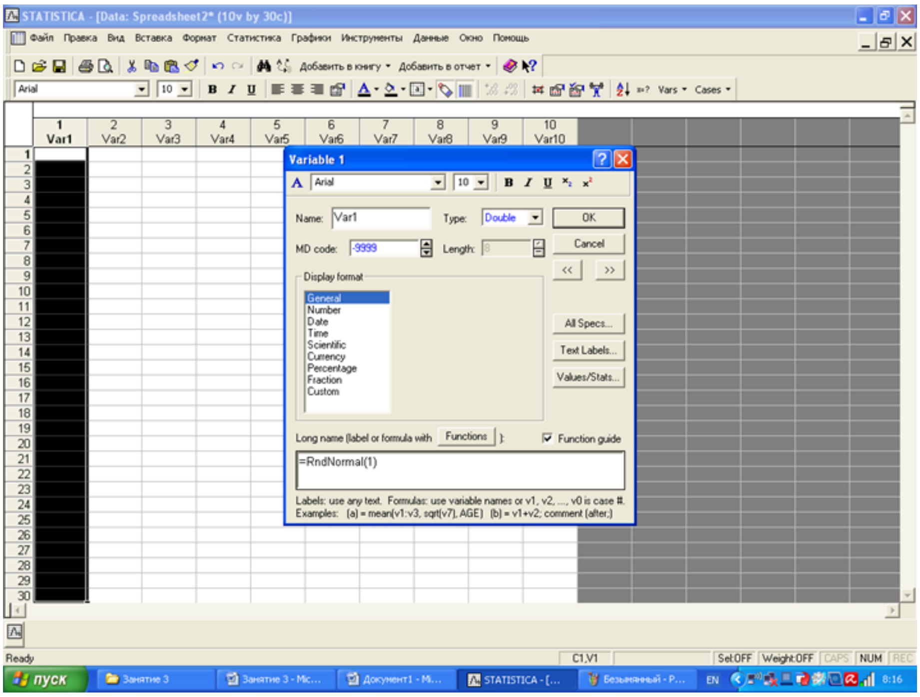


Рис. 31

Постройте гистограмму выборки (рис. 32).

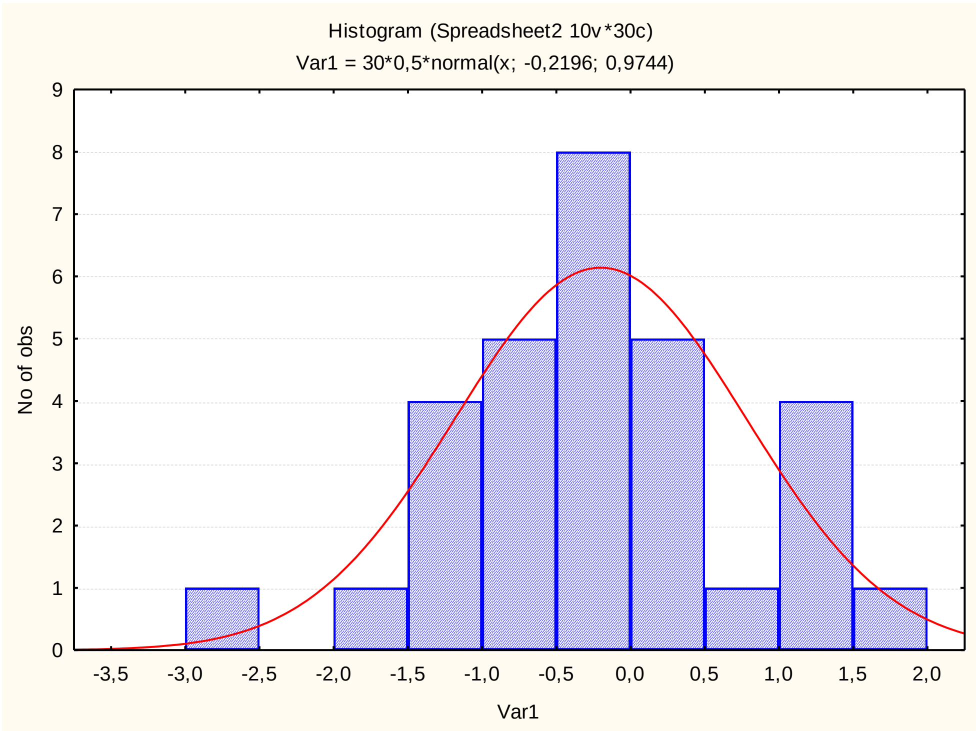


Рис. 32

Переменным второго столбца присвойте значение математического ожидания m = 15 (рис. 33).

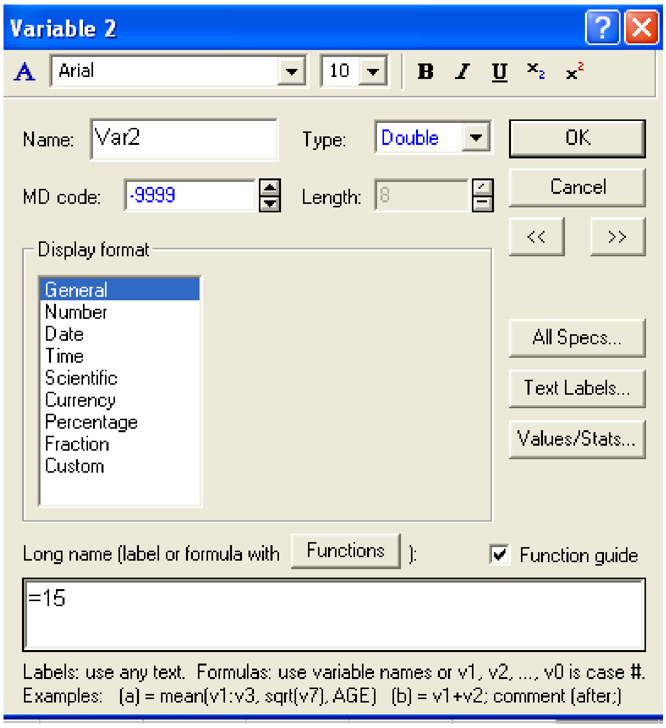


Рис. 33

Переменным третьего столбца присвойте значение среднеквадратичного отклонения $\sigma = 3$ (рис.34).

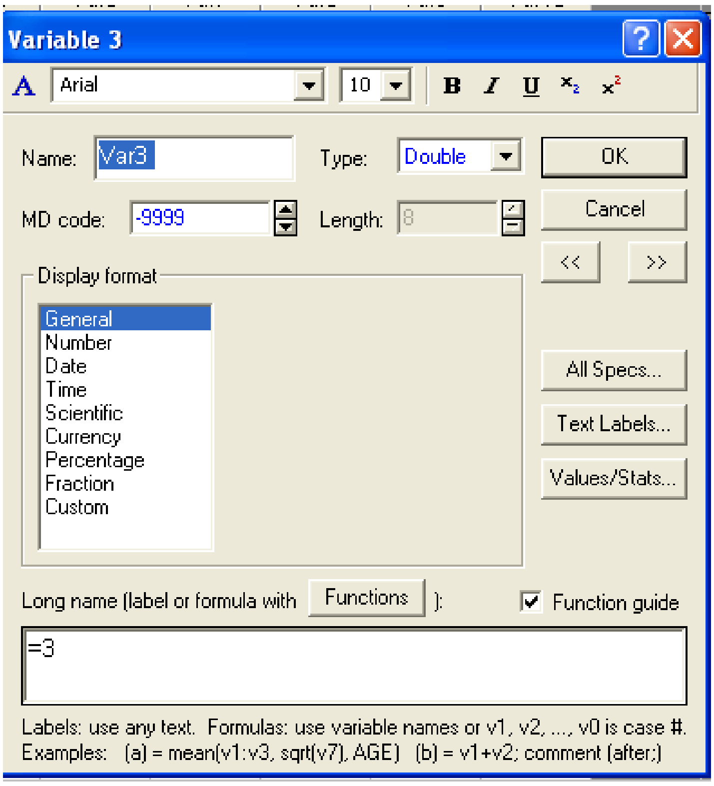


Рис. 34

Переменные (рис. 35) четвертого столбца будут рассчитываться по формуле:

$$x = Z \times \sigma + m \tag{33}$$

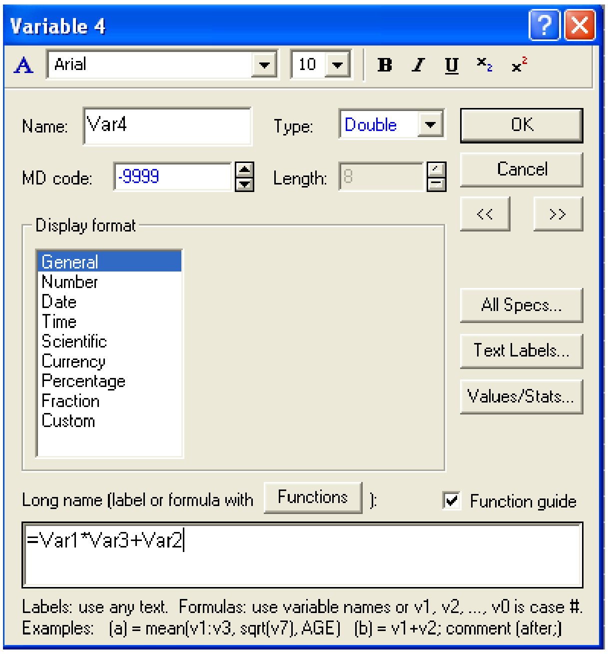


Рис. 35

Постройте гистограмму выборки переменной Var4 (рис. 36).

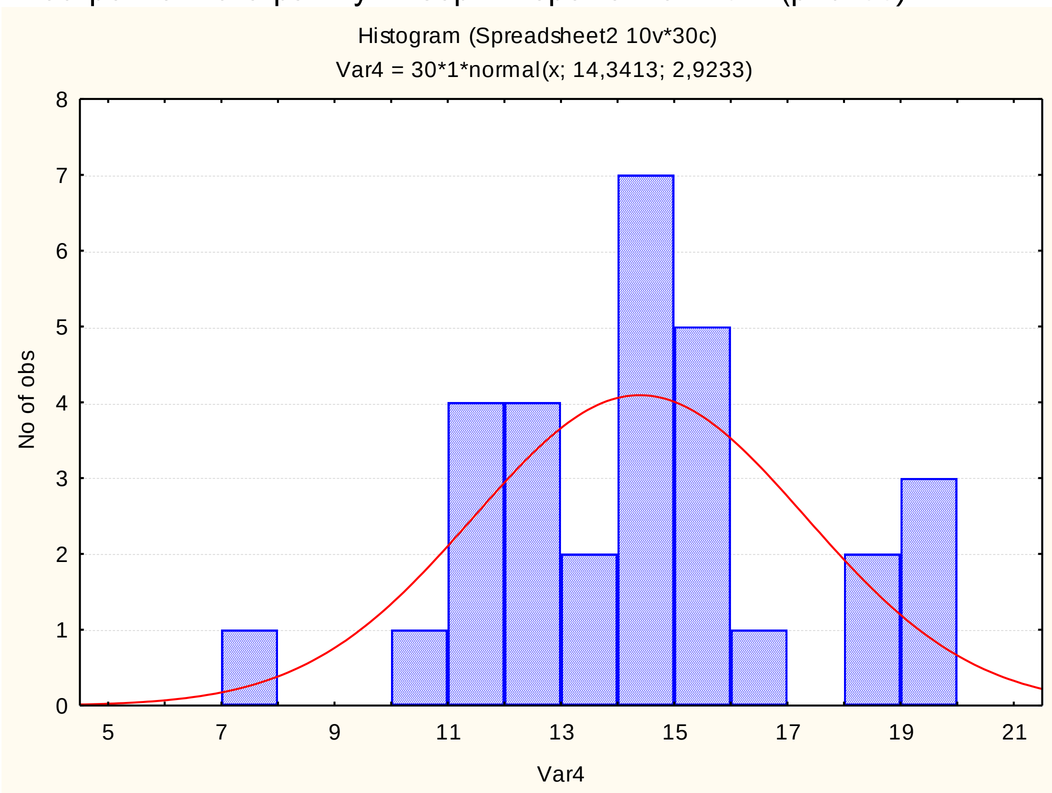


Рис. 36

<u>Задание</u>
В программе Statistica 6.0 произведите генерацию выборки нормального распределения 50 индивидуальных значений - результатов контроля размера деталей с параметрами, указанными в задании.

ПРОВЕРКА ГИПОТЕЗЫ О НОРМАЛЬНОМ РАСПРЕДЕЛЕНИИ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ В STATISTICA 6.0

В программе **Statistica 6.0** выберите в меню **Статистика-Настройка распределения.** В открывшемся диалоговом окне выберите нормальное распределение. Нажмите **ОК** (рис. 37).

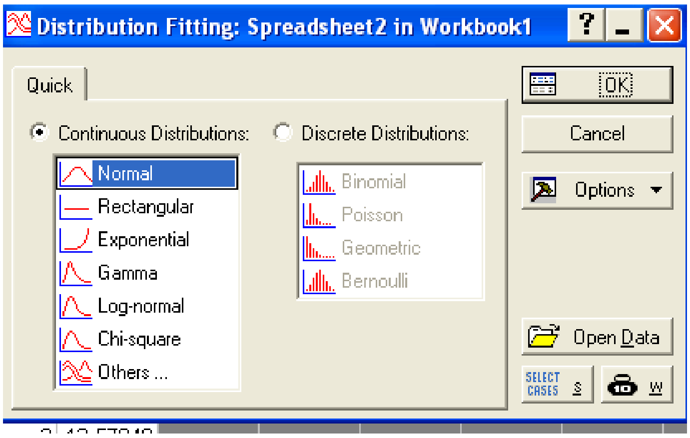


Рис. 37

Задайте число интервалов, равным 7 (рис. 38).

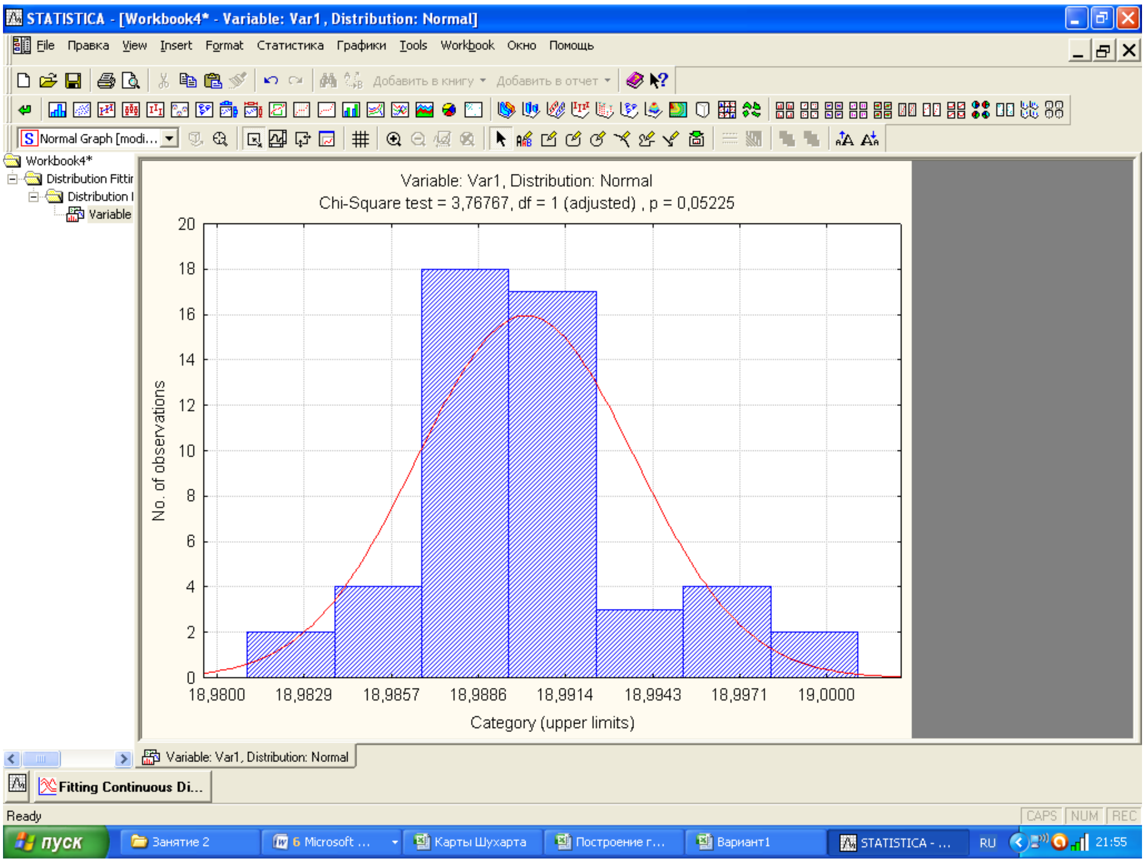


Рис. 38

Если сравнить результаты расчетов с результатами, полученными в MS Excel, то можно увидеть различие. В программе Statistics 6.0 рассматриваемая выборка автоматически разбита на число интервалов, равное 4 (df=1), значение $\chi^2_{набл} = 3{,}76767$.

Вероятность на соответствие нормальному распределению p=5,2%.

<u>Задание</u>
В Statistica 6.0 для 50 индивидуальных значений - результатов контроля размера деталей проверьте гипотезу о нормальном распределении генеральной совокупности по критерию Пирсона.

РАСПРЕДЕЛЕНИЕ ВЫБОРОЧНЫХ СРЕДНИХ

Показатель качества отдельного изделия, взятого из текущего процесса, изменяется от изделия к изделию. Если во время процесса изготовления отбирать выборки, то статистические характеристики этих выборок тоже будут колебаться. Если же по статистическим характеристикам выборок нужно сделать вывод о числовых характеристиках (параметрах) генеральной совокупности, то выборка меньшего объема менее благоприятна для этого, чем выборка большего объема. Но все-таки выгоднее вместо одной большой использовать несколько выборок небольшого объема, которые охватывали бы большой промежуток времени, чтобы получить представление о системе случайных причин, воздействующих на процесс изготовления, т.е. обо всей генеральной совокупности.

Рассмотрим обратную задачу. Пусть известны числовые характеристики генеральной совокупности. Многократно повторяя выборку объема n из генеральной совокупности, вычислим для каждой выборки среднюю арифметическую $\bar{x}$. Так как не все средние арифметические совпадают друг с другом, расположим их в возрастающем порядке. Построив по упорядоченным значениям интервальный ряд, получим распределение средних арифметических значений.

Если количество выборок достаточно велико, то:

1) среднее арифметическое значение $\tilde{x}$ распределения выборочных средних значений совпадает с математическим ожиданием μ- генеральной совокупности;

2) дисперсия распределения выборочных средних $\sigma_{\bar{x}}^2$ зависит от объемов выборок n и связана с дисперсией σ^2 генеральной совокупности следующим соотношением

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}, \qquad (34)$$

или с величиной стандартного отклонения генеральной совокупности

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}, \qquad (35)$$

3) если генеральная совокупность распределена нормально, то распределение средних значений тоже будет нормальным. Но если генеральная совокупность имеет другой закон распределения, то распределение выборочных средних будет приближаться к нормальному распределении, и тем лучше, чем больше объем выборок.

Допустим, что предприятие выполняет заказы на производство партий деталей. Будем считать, что генеральная совокупность состоит из 20000 деталей, из которой отобраны 100 выборок объемом n =5.

С помощью генератора случайных чисел сгенерируйте выборку нормального распределения. Выберите число переменных 5, число случайных чисел – 100, распределение – выберите на вкладке «Нормальное» и сгенерируйте выборку с заданным в варианте средним значением и стандартным отклонением в выходном интервале с A1. Нажмите **ОК.**

В результате получим выборку из 500 значений, соответствующую нормальному распределению.

1) Для каждой из 100 выборок с помощью функции СРЗНАЧ() вычислите среднее арифметическое значение $\bar{x}$.

2) С помощью функции СРЗНАЧ() вычислите μ- генеральной совокупности.

3) С помощью функции СРЗНАЧ() вычислите среднее арифметическое значение $\tilde{x}$ распределения выборочных средних значений.

4) С помощью функции СТАНДОТКЛОН() вычислите стандартное отклонение генеральной совокупности .

5) С помощью функции СТАНДОТКЛОН() вычислите стандартное отклонение распределения выборочных средних значений $\sigma_{\bar{x}}$

6) Убедитесь в справедливости формулы $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

<u>Задание</u>
В MS Office Excel с помощью надстройки Анализ данных для заданного варианта произведите формирование выборки нормального распределения результатов контроля 25 циклов измерений посадочных диаметров партий деталей класса валов (Приложение 2). Задайте число деталей в каждом цикле измерений (число переменных), равным 5, число циклов измерений (число случайных чисел), равным 25. Убедитесь в справедливости формулы $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ ***.***

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ ОЦЕНКИ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ ПРИ ИЗВЕСТНОЙ ДИСПЕРСИИ

Пусть исследуемая случайная величина ***X*** распределена по нормальному закону с известным значением σ и требуется по значению выборочного среднего $\bar{x}_B$ оценить ее математическое ожидание ***m***. Будем рассматривать выборочное среднее $\bar{x}_B$ как случайную величину $\overline{X}$, а значения вариант выборки $\boldsymbol{x}_1$, $\boldsymbol{x}_2$,…, $\boldsymbol{x}_n$ как одинаково распределенные независимые случайные величины $\boldsymbol{X}_1$, $\boldsymbol{X}_2$,…, $\boldsymbol{X}_n$, каждая из которых имеет математическое ожидание ***m*** и среднее квадратическое отклонение σ.

При этом

$$M(\overline{X}) = m, \tag{36}$$

$$\sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}} \tag{37}$$

Оценим вероятность выполнения неравенства $\left|\overline{X} - m\right| < \delta$. Применим формулу для вероятности попадания нормально распределенной случайной величины в заданный интервал:

$$p(\left|\overline{X} - m\right| < \delta) = 2\Phi(\frac{\delta}{\sigma}) \tag{38}$$

Тогда с учетом того, что $\sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}$,

$$p(\left|\overline{X} - m\right| < \delta) = 2\Phi(\frac{\delta\sqrt{n}}{\sigma}) = 2\Phi(t) \tag{39}$$

где $t = \frac{\delta\sqrt{n}}{\sigma}$.

Отсюда

$$\delta = \frac{t\sigma}{\sqrt{n}} \tag{40}$$

И предыдущее равенство можно переписать так:

$$p(\overline{x_b} - \frac{t\sigma}{\sqrt{n}} < m < \overline{x_b} + \frac{t\sigma}{\sqrt{n}}) = 2\Phi(t) = \gamma \tag{41}$$

Итак, значение математического ожидания *m* с вероятностью (надежностью) γ попадает в интервал $\left(\bar{x}_B - \frac{t\sigma}{\sqrt{n}}; \bar{x}_B + \frac{t\sigma}{\sqrt{n}}\right)$, где значение *t* определяется из таблиц для функции Лапласа так, чтобы выполнялось равенство 2Φ(*t*) = γ.

ТОЧЕЧНАЯ И ИНТЕРВАЛЬНАЯ ОЦЕНКА СРЕДНЕГО ЗНАЧЕНИЯ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ ПРИ ИЗВЕСТНОЙ ДИСПЕРСИИ ПО АЛГОРИТМУ ГОСТ Р 50779.21-96

В таблице 5 представлены результаты расчетов для сгенерированных ранее 50 индивидуальных значений - результатов контроля размера деталей.

α - риск первого рода (вероятность отвергнуть гипотезу, когда она верна);

(1 - α) - доверительная вероятность, где α, $0 < \alpha < 1$, - уровень значимости при проверке гипотез;

$u_{1-\alpha}$, $u_{1-\alpha/2}$ - квантили стандартного нормального закона распределения уровней 1 - α и 1 - α/2 соответственно.

Таблица 5

<table>
<tr><td>Статистические и исходные данные</td><td>Табличные данные и вычисления</td></tr>
<tr><td>1 Объем выборки:
n =50

2 Сумма значений наблюдаемых величин:
Σx =949,507
3 Известное значение дисперсии:
σ^2_0=0,003^2

4 Выбранная доверительная вероятность:
1 - α=0,95</td><td>1 Квантиль стандартного нормального закона распределения уровня (1 - α):
$u_{1-\alpha}$ =1,65
2 Квантиль стандартного нормального закона распределения уровня (1 - α/2):
$u_{1-\alpha/2}$ =1,96
3 Вычисляем:
$\bar{x} = \frac{1}{n}\sum x$
$\bar{x} = 18{,}990$
4 Вычисляем:
$K_1 = \frac{u_{1-\alpha}}{\sqrt{n}}$
$K_1 = \frac{1{,}65}{\sqrt{50}} = 0{,}233$
5 Вычисляем:
$K_2 = \frac{u_{1-\alpha/2}}{\sqrt{n}}$
$K_1 = \frac{1{,}96}{\sqrt{50}} = 0{,}277$</td></tr>
<tr><td colspan="2">Результаты:
1 Точечная оценка параметра μ:
$\hat{\mu} = \bar{x}$ =18,990
2 Двусторонний симметричный доверительный интервал для μ:
$\bar{x} - K_2\sigma_0 \le \mu \le \bar{x} + K_2\sigma_0$
$18{,}9893 \le \mu \le 18{,}9910$
3 Односторонние доверительные интервалы для μ:
$\mu \le \bar{x} + K_1\sigma_0$= 18,9908 или
$\mu \ge \bar{x} - K_1\sigma_0$=18,9894</td></tr>
</table>

Задание

Произведите точечную и интервальную оценку среднего значения нормального распределения при известной дисперсии для 50 индивидуальных значений - результатов контроля размера деталей по алгоритму ГОСТ Р 50779.21-96.

РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА

Рассмотрим набор результатов x_1, x_2,.....,x_n многократного измерения нормально распределенной величины x . Из этих данных получены оценки $\bar{x}$ и $\sigma_{\bar{x}}$. Проверяется гипотеза о том, что $\bar{x}=x_0$, где x_0 – заданное значение измеряемой величины, точно известное, например, из расчетов или справочных таблиц.

Введем новую величину, содержащую как экспериментальное среднее, так и заданное значение:

$$t = \frac{\bar{x} - x_0}{\sigma_{\bar{x}}} \tag{42}$$

Если равенство $\bar{x}= x_0$ справедливо для $n \to \infty$, то распределение величины t при конечном количестве измерений n будет распределением Стьюдента.

Плотность вероятности распределения Стьюдента описывает выражение

$$\rho(t,n) = \frac{\Gamma\left(\frac{n}{2}\right)\cdot\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}}}{\sqrt{\pi(n-1)}\cdot\Gamma\left(\frac{n-1}{2}\right)} \tag{43}$$

$$\Gamma(m) = \int_0^\infty e^{-y} y^{m-1} dy$$

где n – количество проведенных измерений, а m>0.

Зная $\rho(t,n)$, не составит труда вычислить интервал, в который величина t попадет с заданным уровнем доверия. Для этого необходимо решить уравнение

$$\int_{-t(\alpha,n)}^{+t(\alpha,n)} \rho(t,n)dt = \alpha$$

Вероятность α определяет так называемый **уровень значимости**.

Если значение $t = \frac{\bar{x} - x_0}{\sigma_{\bar{x}}}$ попадает в указанный интервал, то это свидетельствует в пользу справедливости гипотезы о совпадении $\bar{x}$ и x_0 при уровне значимости α . Чем больше α , тем шире интервал, тем больше вероятность обнаружить в нем величину t , относящуюся к эксперименту при $\bar{x} = x_0$.

Найдем интервал возможного изменения величины $\bar{x}$. Воспользуемся

$$-t(\alpha,n)\le\frac{\bar{x}-x_0}{\sigma_{\bar{x}}}\le+t(\alpha,n) \quad (44)$$

откуда

$$\bar{x}-t(\alpha,n)\sigma_{\bar{x}}\le x_0\le\bar{x}+t(\alpha,n)\sigma_{\bar{x}} \quad (45)$$

Если при сравнении $\bar{x}$ и x_0 значение x_0 попадает в доверительный интервал вокруг $\bar{x}$ то статистическим выводом является заключение о совпадении сравниваемых величин с доверительной вероятностью α.

В измерениях принято использовать вероятность α =0,68, в пределе при больших n задающую интервал $\pm\sigma_{\bar{x}}$ вокруг $\bar{x}$.

Для повышения достоверности сравнения используют уровень значимости α =0,997, определяющий более широкий интервал, в пределе стремящийся к $\pm3\sigma_{\bar{x}}$.

Для малых n за погрешность прямого многократного измерения величины x естественно принимать $\Delta x=t(\alpha,n)\sigma_{\bar{x}}$. Именно в этом интервале могут оказаться точные величины x_0, совпадающие с результатом измерения $\bar{x}$.

ТОЧЕЧНАЯ И ИНТЕРВАЛЬНАЯ ОЦЕНКА СРЕДНЕГО ЗНАЧЕНИЯ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ ПРИ НЕИЗВЕСТНОЙ ДИСПЕРСИИ ПО АЛГОРИТМУ ГОСТ Р 50779.21-96

В таблице 6 представлены результаты расчетов для сгенерированных ранее 50 индивидуальных значений - результатов контроля размера деталей.

Таблица 6

Статистические и исходные данные	Табличные данные и вычисления
1 Объем выборки: n =50	1 Квантиль распределения Стьюдента уровня (1 - α) с v степенями свободы: $t_{1-\alpha}(v)$ =1,676
2 Сумма значений наблюдаемых величин: Σx =949,507	2 Квантиль распределения Стьюдента уровня (1 - α/2) с v степенями свободы: $t_{1-\alpha/2}(v)$ =2,010
3 Сумма квадратов значений наблюдаемых величин: Σx^2 =18031,268	3 Вычисляем: $\bar{x} = \frac{\sum x}{n}$ $\bar{x} = 18{,}990$
4 Степени свободы: $v = n - 1$ =49	4 Вычисляем: $\frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum x^2 - (\sum x)^2/n}{n-1}$ =0,000013
5 Выбранная доверительная вероятность: 1-α= 0,95	5 Вычисляем: $S = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$ =0,0035
	6 Вычисляем: $l_1 = \frac{t_{1-\alpha}(v)}{\sqrt{n}}$ $l_1 = \frac{1{,}676}{\sqrt{50}}$ =0,2370
	7 Вычисляем: $l_2 = \frac{t_{1-\alpha/2}(v)}{\sqrt{n}}$ $l_2 = \frac{2{,}010}{\sqrt{50}} = 0{,}2843$

Статистические и исходные данные	Табличные данные и вычисления
Результаты: 1 Точечная оценка параметра μ: $\hat{\mu} = \bar{x}$ =18,990 2 Точечная оценка параметра D: $\hat{D} = S^2$ $\hat{D}$ =0,000013 3 Двусторонний симметричный доверительный интервал для параметра μ: $\bar{x} - l_2 S \le \mu \le \bar{x} + l_2 S$ $18{,}9890 \le \mu \le 18{,}9910$ 4 Односторонние доверительные интервалы для параметра μ: $\mu \le \bar{x} + l_1 S$= 18,9908 или (1) $\mu \ge \bar{x} - l_1 S$=18,9892 (2)	

Задание

Произведите точечную и интервальную оценку среднего значения нормального распределения при неизвестной дисперсии 50 индивидуальных значений - результатов контроля размера деталей по алгоритму ГОСТ Р 50779.21-96.

ТОЧЕЧНАЯ И ИНТЕРВАЛЬНАЯ ОЦЕНКА ДИСПЕРСИИ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ ПО АЛГОРИТМУ ГОСТ Р 50779.21-96

В таблице 7 представлены результаты расчетов для сгенерированных ранее 50 индивидуальных значений - результатов контроля размера деталей.

Таблица 7

Статистические и исходные данные	Табличные данные и вычисления
1 Объем выборки: $n = 50$ 2 Сумма значений наблюдаемых величин: $\Sigma x = 949{,}507$ 3 Сумма квадратов значений наблюдаемых величин: $\Sigma x^2 = 18031{,}268$ 4 Степени свободы: $v = n - 1 = 49$ 5 Выбранная доверительная вероятность: $1 - \alpha = 0{,}95$	1 Квантили χ^2-распределения с v степенями свободы уровней α, $(1 - \alpha)$, $\alpha/2$ и $(1 - \alpha/2)$ соответственно $\chi^2_{\alpha}(v) = 33{,}93$ $\chi^2_{1-\alpha}(v) = 66{,}34$ $\chi^2_{\alpha/2}(v) = 31{,}55$ $\chi^2_{1-\alpha/2}(v) = 70{,}22$ 2 Вычисляем: $\sum (x-\bar{x})^2 = \sum x^2 - (\sum x)^2 / n =$ $= 0{,}000625$ 3 Вычисляем: $S^2 = \dfrac{\sum (x-\bar{x})^2}{n-1}$ $=0{,}000011$

Результаты:
1 Точечные оценки дисперсии D и стандартного отклонения σ генеральной совокупности: $\widehat{D} = S^2$ $\widehat{D} = 0{,}000013$ $\widehat{\sigma} = \sqrt{S^2}$ $=0{,}0036$ 2 Двусторонний доверительный интервал* для дисперсии D: $\dfrac{\sum (x-\bar{x})^2}{\chi^2_{1-\alpha/2}(v)} < D < \dfrac{\sum (x-\bar{x})^2}{\chi^2_{1/\alpha}(v)}$ $\dfrac{0{,}000625}{70{,}22} < D < \dfrac{0{,}000625}{31{,}55}$ 0,0000089<D<0,0000198 3 Односторонний доверительный интервал* для дисперсии D: $D > \dfrac{\sum (x-\bar{x})^2}{\chi^2_{1-\alpha}(v)} = \sigma^2_{н}$ (3) $D > \dfrac{0{,}000625}{66{,}34}$ $D > 0{,}0000098$ или $D < \dfrac{\sum (x-\bar{x})^2}{\chi^2_{\alpha}(v)} = \sigma^2_{в}$ (4)

Статистические и исходные данные	Табличные данные и вычисления
$D > \frac{0{,}000625}{33{,}93}$ *D*<0,0000192	
* *Значения границ доверительного интервала стандартного отклонения* σ *являются корнем квадратным из значений границ доверительного интервала дисперсии D.*	

<u>Задание</u>

Произведите точечную и интервальную оценку дисперсии нормального распределения для 50 индивидуальных значений - результатов контроля размера деталей по алгоритму ГОСТ Р 50779.21-96.

ПОДСЧЕТ ВЕРОЯТНОСТЕЙ РАСПРЕДЕЛЕНИЙ В STATISTICA 6.0

Откройте программу **Statistica 6.0**. В программе можно вычислить данные таблиц распределений, приведенных в приложениях А, Б, В. Для этого выделите в меню пиктограмму **Статистика,** в раскрывшемся списке **Подсчет вероятности-Распределения** (рис. 39). Для определения значений границ среднего значения при известной дисперсии выберите распределение Z (Normal), задайте среднее значение, стандартное отклонение, значение вероятности. Рассчитайте левую границу одностороннего доверительного интервала среднего значения. При выборе параметра (1-совокупный р) рассчитайте правую границу одностороннего доверительного интервала среднего значения. При выборе параметра Двойной критерий рассчитаны границы двустороннего симметричного доверительного интервала среднего значения.

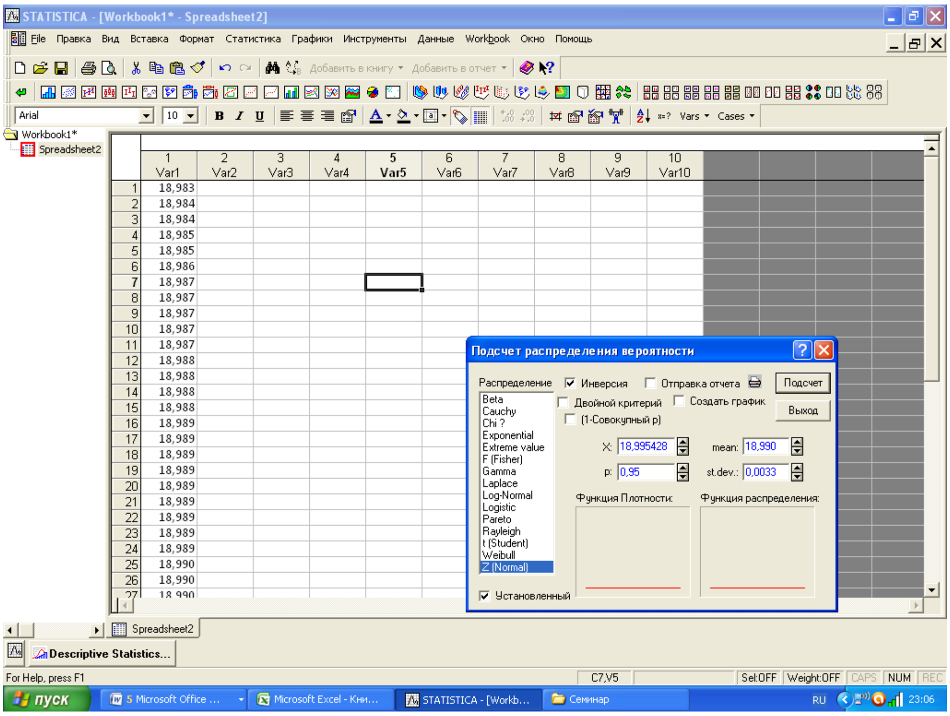


Рис. 39

Значения параметров других распределений, приведенных в ГОСТ Р 50779.21-96, рассчитываются аналогично.

КОНТРОЛЬНЫЕ КАРТЫ

Стабильный производственный процесс - процесс, в котором отсутствует системная вариация, то есть процесс, который последовательно производит продукцию, обладающую одинаковыми свойствами. Мы можем проверить любую партию продукции или любой период наработки продукции на линии (станке) на стабильность процесса. Стабильность означает заранее известные результаты, что гарантирует нам качество.

Предположите, что вы получили заказ на изготовление партии деталей в количестве 50 штук с контролируемым размером $19h7_{-0,021}$(данные сгенерированной ранее выборки).

Постройте линейчатую диаграмму распределения эмпирических частот (рис. 40).

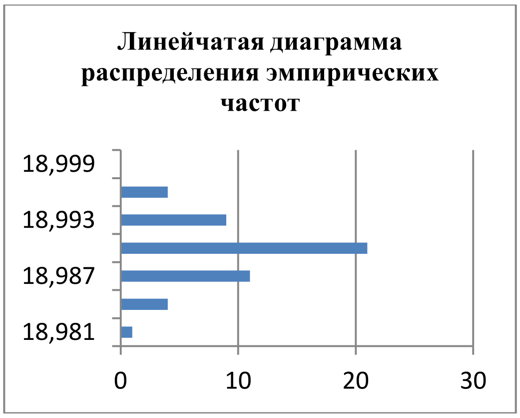


Рис. 40

Методика построения контрольных карт по количественному признаку

Предположите, что результаты измерений контролируемого размера вы фиксируете по ходу изготовления деталей и данные измерений наносите на график. Поскольку предыдущая выборка была упорядочена по возрастанию для построения гистограммы распределения сгруппированных данных, откройте новую книгу MS Excel и в выходном интервале A1 сгенерируйте выборку нормального распределения, аналогичную предыдущей, с числом переменных 1, числом случайных чисел – 50, с параметрами, рассчитанными по алгоритму ГОСТ Р 50779.21-96: Среднее значение = 18,990, Стандартное отклонение = 0,0036.По полученным данным постройте график (рис. 41).

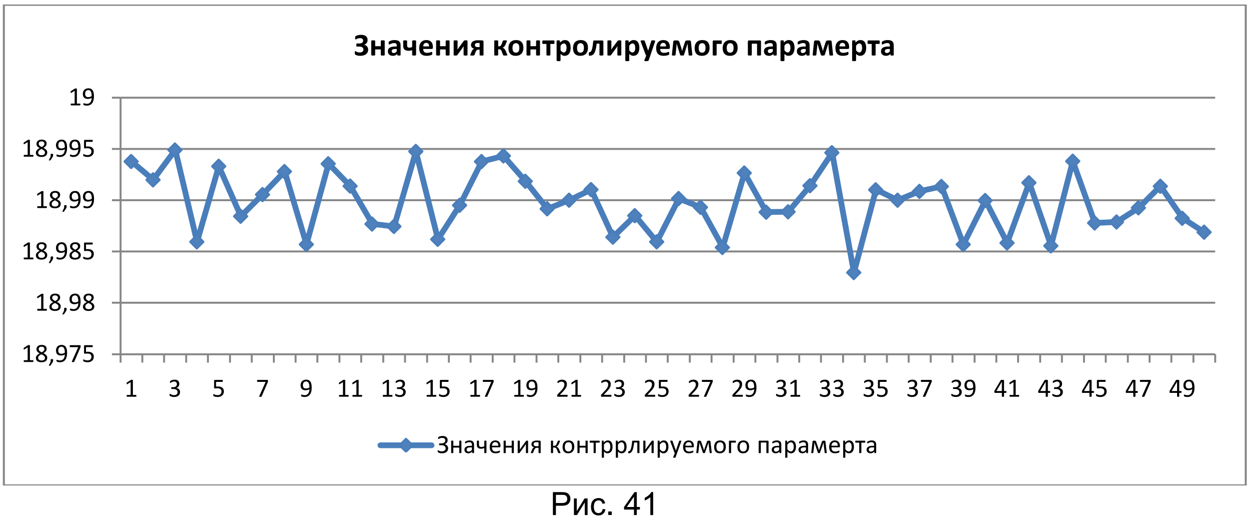


Рис. 41

По сути это та же линейчатая диаграмма распределения эмпирических частот, только показанная другим инструментом.

В случае если процесс стабилен, то вероятность появления детали с размерами нижнего поля допуска LTL(НГД)= 18,979 и верхнего поля допуска UTL(ВГД)=19,000 крайне мала. Если появилась деталь с таким размером, – её появление не случайно. Какая - то внешняя причина вызвала это появление. Требуется срочно проверить станок, персонал, сырье и т.д.

Это меняет привычные представления. До сих пор считалось, что проблема существует тогда, когда деталь не соответствует требованиям чертежа.

Любую группу данных можно охарактеризовать средним значением (сумма всех значений, поделенная на количество значений) и среднеквадратическим отклонением (стандартное отклонение, показатель, который характеризует рассеивание значений в группе, степень близости значений к среднему группы).

1 Большое стандартное отклонение, большой разброс между максимальными значениями, минимальными и средним в группе.

2 Маленькое стандартное отклонение, небольшой разброс между максимальными значениями, минимальными и средним в группе.

Среднеквадратичное отклонение (далее сигма) имеет большое прикладное применение в разных процедурах контроля качества.

В стабильном процессе с нормальным распределением значения располагаются на определенном удалении от среднего.

Вероятность того, что каждая новая нарабатываемая деталь в стабильном процессе будет иметь размер в пределах одной сигма от среднего значения – 68,2 %.

Вероятность того, что каждая новая нарабатываемая деталь в стабильном процессе будет иметь размер в пределах двух сигм от среднего значения – 95,5 %.

Вероятность того, что каждая новая нарабатываемая деталь в стабильном процессе будет иметь размер в пределах трех сигм от среднего значения – 99,7 %.

Положение верхней контрольной границы ВКГ(UCL) и нижней контрольно границы НКГ(LCL) определяется аналитически либо по специальным таблицам и зависит от объёма выборки. UCL и LCL служат для предупреждения разладки процесса, когда изделия еще соответствуют техническим требованиям.

При достаточно большом объеме выборки пределы UCL и LCL определяют по формулам

$$\text{UCL(ВКГ)} = \bar{x} + 3\sigma \ \text{(UCL)} \tag{46}$$

$$\text{LCL(НКГ)} = \bar{x} - 3\sigma \ \text{(LCL)} \tag{47}$$

Для полученной выборки определите $\bar{x}$, σ, UCL,LCL

Дисперсия генеральной совокупности связана со значением выборочной дисперсии следующим соотношением:

$$\sigma^2 \approx \frac{n}{n-1} \times s^2 \tag{48}$$

В таблице представлены параметры процесса изготовления партии деталей с контролируемым размером $19h7_{-0,021}$

Таблица 8

Параметры процесса	Значения
Среднее значение	18,9898
Выборочная дисперсия	0,0000089
Дисперсия генеральной совокупности	0,0000091
Стандартное отклонение	0,003
UCL (ВКГ)	18,9988
LCL (НКГ)	18,9808
UTL (ВГД)	19,0000
LTL (НГД)	18,9790

Добавьте на карту линии UCL, LCL, Ср знач ($\bar{x}$, UTL, LTL (рис. 42).

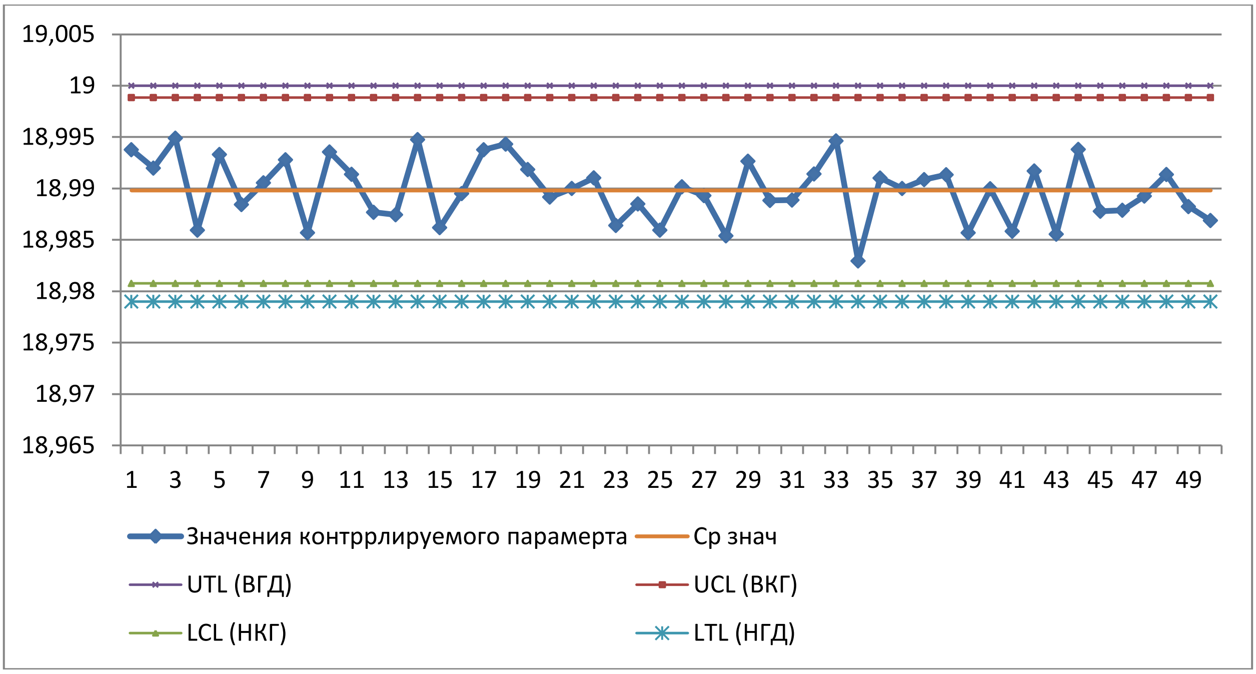


Рис. 42

Как видно по графику, все результаты измерений не вышли за контрольные границы, следовательно, разладка технологического процесса отсутствует и его регулирование не требуется.

Контрольные карты применяются, когда требуется установить характер неисправностей и дать оценку стабильности процесса; когда необходимо установить, нуждается ли процесс в регулировании или его необходимо оставить таким, каков он есть.

Контрольной картой можно также подтвердить улучшение процесса.

Контрольная карта является средством распознания отклонений из-за неслучайных или особых причин от вероятных изменений, присущих процессу. Вероятные изменения редко повторяются в прогнозируемых пределах. Отклонения из-за неслучайных или особых причин сигнализируют о том, что некоторые факторы, влияющие на процесс, необходимо идентифицировать, расследовать и поставить под контроль.

Информация о контрольных картах содержится в стандартах ГОСТ Р 50779.42-99 (ИСО 8258-91) и ГОСТ Р 51814.3-2001.

Наибольшее распространение получили контрольные карты среднего значения X и контрольные карты размаха R, которые используются совместно или раздельно. Контролироваться должны естественные колебания между пределами контроля. Нужно убедиться, что выбран правильный тип контрольной карты для определенного типа данных. Данные должны быть взяты точно в той последовательности, в какой собраны, иначе они теряют смысл. Не следует вносить изменения в процесс в период сбора данных. Данные должны отражать, как процесс идет естественным образом.

Контрольная карта может указать на наличие потенциальных проблем до того, как начнется выпуск дефектной продукции.

Принято говорить, что процесс вышел из-под контроля, если одна или более точек вышли за пределы контроля.

Существуют два основных типа контрольных карт: для качественных (годен – негоден) и для количественных признаков. Для качественных признаков возможны четыре вида контрольных карт: число дефектов на единицу продукции; число дефектов в выборке; доля дефектных изделий в выборке; число дефектных изделий в выборке. При этом в первом и третьем случаях объем выборки будет переменным, а во втором и четвертом – постоянным.

Таким образом, целями применения контрольных карт могут быть:

- выявление неуправляемого процесса;
- контроль за управляемым процессом;
- оценивание возможностей процесса.

Обычно подлежит изучению следующая переменная величина (параметр процесса) или характеристика:

- известная важная или важнейшая;
- предположительная ненадежная;
- по которой нужно получить информацию о возможностях процесса;
- эксплуатационная, имеющая значение при маркетинге.

При этом не следует контролировать все величины одновременно. Ведение контрольных карт требует затрат, поэтому нужно использовать их разумно: тщательно выбирать характеристики; прекращать работу с картами при достижении цели: продолжать вести карты только тогда, когда процессы и технические требования сдерживают друг друга.

Необходимо иметь в виду, что процесс может быть в состоянии статистического регулирования и давать 100% брака. И наоборот, может быть неуправляемым и давать продукцию, на 100% отвечающую техническим требованиям.

Контрольные карты позволяют проводить анализ возможностей процесса. Возможности процесса – это способность функционировать должным образом. Как правило, **под возможностями процесса понимают способность удовлетворять техническим требованиям.**

Виды контрольных карт

Контрольные карты для регулирования по количественным признакам (измеренные величины выражаются количественными значениями):

1) Контрольная карта состоит из контрольной карты, отражающей контроль за изменением среднего арифметического, и контрольной карты R, служащей для контроля изменений рассеивания значений показателей

качества. Применяется при измерении таких показателей, как длина, масса, диаметр, время, предел прочности при растяжении, шероховатость, прибыль и т.д.

2) Контрольная карта состоит из контрольной карты, осуществляющей контроль за изменением значения медианы, и контрольной карты R. Применяется в тех же случаях, что и предыдущая карта. Однако она более проста, поэтому более пригодна для заполнения на рабочем месте.

3) Контрольные карты для регулирования по качественным признакам:

4) Контрольная карта p (для доли дефектных изделий) или процента брака, применяется для контроля и регулирования технологического процесса после проверки небольшой партии изделий и разделения их на доброкачественные и дефектные, т.е. определения их по качественным признакам. Доля дефектных изделий получена путём деления числа обнаруженных дефектных изделий на число проверенных изделий. Может применяться также для определения интенсивности выпуска продукции, процента неявки на работу и т.д.

5) Контрольная карта pn (количество брака), применяется в случаях, когда контролируемым параметром является число дефектных изделий при постоянном объеме выборки n. Практически совпадает с картой p;

6) Контрольная карта c (число дефектов на одно изделие), используется, когда контролируется число дефектов, обнаруживаемых среди постоянных объемов продукции (автомобили – одна или 5 транспортных единиц, листовая сталь – один или 10 листов).

7) Контрольная карта n (число дефектов на единицу площади), используется, когда площадь, длина, масса, объём, сорт непостоянны и обращаться с выборкой как с постоянным объемом невозможно.

Виды индексов для оценки процессов

Для определения воспроизводимости (пригодности) процесса используют индексы *Cp (Pp), CR (PR), Cpк (Ppк), CPL (PPL), CPU (PPU).* Индексы *Cp, Cpк, CR, CPL, CPU* используются в случаях, когда исследуемый процесс статистически управляем, в противном случае используются индексы *Pp, Ppк, PR, PPL, PPU*.

Индекс воспроизводимости процесса *Cp* определяется как соотношение ширины поля допуска к ширине 6s (99,7%) зоны рассеивания статистически контролируемого процесса: где LTL— нижняя граница поля допуска, UTL— верхняя граница поля допуска. Значение параметра не известно, а имеется только его статистическая оценка — выборочное стандартное отклонение s.

Дисперсия генеральной совокупности связана со значением выборочной дисперсии следующим соотношением (48)

$$\sigma^2 \approx \frac{n}{n-1} \times s^2$$

Если известно истинное значение отклонение его необходимо использовать вместо s.

Для анализа можно использовать следующие ограничения: C<1 — неудовлетворительно; C>1,00 — удовлетворительно; C >1,33—хорошо.

Индекс воспроизводимости процесса *Cp* рассчитывается по формуле:

$$C_p(P_p) = \frac{UTL - LCL}{6\sigma} \qquad (49)$$

где UTL – верхняя граница допуска;
LTL – нижняя граница поля допуска;
σ –стандартное отклонение.

Индекс воспроизводимости *Cpk* характеризует рассеивание с учетом настроенности процесса на центр поля допуска.

Индекс равен отношению разности между средним процесса и ближайшим пределом поля допуска к половине присущей стабильному процессу изменчивости: где положение центра рассеивания, на практике принимаемое равным среднему .

Показатель *Cpк* близок к *Cp,* но использует среднее процесса и может рассматриваться как показатель его работоспособности. Определить значение *Cpк* можно по формуле:

$$C_{pk}(P_{pk}) = \min[PPU; PPL] \qquad (50)$$

Показатели *PPU* и *PPL* являются верхним и нижним показателем пригодности процесса. Индексы *PPU* и *PPL* вычисляются по формулам:

$$s^2 = \frac{\Delta}{n}\sum_{j=1}^{m}\frac{(x_j - \bar{x})^2}{\Delta_j}n_j \quad PPU = \frac{UTL - \bar{x}}{3\sigma} \qquad (51)$$

$$PPL = \frac{\bar{x} - LTL}{3\sigma} \qquad (52)$$

где UTL - верхняя граница допуска;
LTL- нижняя граница поля допуска;
- среднее значение;
–стандартное отклонение.

Показатель *Cp* описывает теоретическое состояние процесса (без смещения относительно центра поля допуска), *Cpк* – фактическое. *Cpk<Cp.*

Одними из основных показателей возможностей процессов, наряду с уже рассмотренными индексами воспроизводимости *Cp* и Cpk являются также **индексы пригодности *Pp* и *Ppk*.** Индексы *Pp и Ppk* определяются аналогично *Cp* и *Cpk*, отличие заключается лишь в том, что оценка параметров вычисляется на основе вариации процесса в течение значимого производственного цикла, а не «мгновенного» рассеивания, то есть среднего рассеивания внутри одной пробы.

Коэффициент возможностей процесса CR – коэффициент воспроизводимости процесса, применяемый для стабильных процессов, представляет собой величину, обратную индексу C_p, то есть:

$$CR = \frac{1}{C_p} \qquad (53)$$

При *CR ≤ 0,75* технологический процесс достаточно точный.

При *CR= 0,76…0,98* технологический процесс требует внимательного наблюдения.

При *CR> 0,98* точность технологического процесса неудовлетворительная.

Коэффициент пригодности процесса PR – коэффициент пригодности процесса, стабильность которого не подтверждена, представляет собой величину, обратную индексу P_p, то есть:

$$PR = \frac{1}{P_p} \tag{54}$$

В таблице 9 представлены результаты расчета индексов процесса изготовления партии деталей в количестве 50 штук с контролируемым размером $19h7_{-0,021}$

Таблица 9

Расчет индексов процесса	Значения
Среднее значение	18,9898
Выборочная дисперсия	0,0000089
Дисперсия генеральной совокупности	0,0000091
Стандартное отклонение	0,0030
UCL (ВКГ)	18,9988
LCL (НКГ)	18,9808
UTL (ВГД)	19,0000
LTL (НГД)	18,9790
Индекс воспроизводимости *Cp*	1,16
Верхний показатель пригодности *PPU*	1,13
Верхний показатель пригодности *PPL*	1,20
Индекс воспроизводимости *Cpk*	1,13
Коэффициент возможностей процесса *CR*	0,86
Коэффициент пригодности процесса *PR*	0,89

ГОСТ Р 50779.42-99 (ИСО 8258-91) КОНТРОЛЬНЫЕ КАРТЫ ШУХАРТА

Контрольная карта - это графическое средство, использующее статистические подходы, важность которых для управления производственными процессами была впервые показана доктором У.Шухартом в 1924 г.

Карта Шухарта требует данных, получаемых выборочно из процесса через примерно равные интервалы. Интервалы могут быть заданы либо по времени (например, ежечасно), либо по количеству продукции (каждая партия). Обычно каждая подгруппа состоит из однотипных единиц продукции или услуг с одними и теми же контролируемыми показателями, и все подгруппы имеют равные объемы. Для каждой подгруппы определяют одну или несколько характеристик, таких как среднее арифметическое подгруппы $\overline{X}$ и размах подгруппы R или выборочное стандартное отклонение $\overline{\overline{S}}$. Карта Шухарта - это график значений определенных характеристик подгрупп в зависимости от их номеров. Она имеет центральную линию (CL), соответствующую эталонному значению характеристики. При оценке того, находится ли процесс в статистически управляемом состоянии, эталонным обычно служит среднее арифметическое рассматриваемых данных. При управлении процессом эталонным служит долговременное значение характеристики, установленное в технических условиях, или ее номинальное значение, основанное на предыдущей информации о процессе, или намеченное целевое значение характеристики продукции или услуги. Карта Шухарта имеет две статистические определяемые контрольные границы относительно центральной линии, которые называются верхней контрольной границей (UCL) и нижней контрольной границей (LCL).

Контрольные границы на карте Шухарта находятся на расстоянии 3σ от центральной линии, где σ - генеральное стандартное отклонение используемой статистики. Изменчивость внутри подгрупп является мерой случайных вариаций. Для получения оценки σ вычисляют выборочное стандартное отклонение или умножают выборочный размах на соответствующий коэффициент. Эта мера не включает межгрупповых вариаций, а оценивает только изменчивость внутри подгрупп.

Границы $\pm 3\sigma$ указывают, что около 99,7% значений характеристики подгрупп попадут в эти пределы при условии, что процесс находится в статистически управляемом состоянии. Другими словами, есть риск, равный 0,3%, что нанесенная точка окажется вне контрольных границ, когда процесс стабилен.

Вероятность того, что нарушение границ в самом деле случайное событие, а не реальный сигнал, считается столь малой, что при появлении точки вне границ следует предпринять определенные действия. Так как действие предпринимается именно в этой точке, то 3σ контрольные границы иногда называются "границами действий".

Часто на контрольной карте границы проводят еще и на расстоянии 2σ. Тогда любое выборочное значение, попадающее за границы 2σ, может служить предостережением о грозящей ситуации выхода процесса из состояния статистической управляемости. Поэтому границы $\pm 2\sigma$ иногда называют "предупреждающими".

При применении контрольных карт возможны два вида ошибок: первого и второго рода.

Ошибка первого рода возникает, когда процесс находится в статистически управляемом состоянии, а точка выскакивает за контрольные границы случайно. В результате неправильно решают, что процесс вышел из состояния статистической управляемости, и делают попытку найти и устранить причину несуществующей проблемы.

Ошибка второго рода возникает, когда рассматриваемый процесс не управляем, а точки случайно оказываются внутри контрольных границ. В этом случае неверно заключают, что процесс статистически управляем и упускают возможность предупредить рост выхода несоответствующей продукции. Риск ошибки второго рода - функция трех факторов: ширины контрольных границ, степени неуправляемости и объема выборки. Их природа такова, что можно сделать лишь общее утверждение о величине ошибки.

Система карт Шухарта учитывает только ошибки первого рода, равные 0,3% в пределах границ 3σ.

Если процесс статистически управляем, контрольные карты реализуют метод непрерывной статистической проверки нулевой гипотезы о том, что процесс не изменился и остается стабильным.

Когда наносимое значение выходит за любую из контрольных границ или серия значений проявляет необычные структуры, состояние статистической управляемости подвергается сомнению. В этом случае надо исследовать и обнаружить неслучайные (особые) причины, а процесс можно остановить или скорректировать. Как только особые причины найдены и исключены, процесс снова готов к продолжению работы. При возникновении ошибки первого рода можно не найти никакой особой причины. Тогда считают, что выход точки за границы представляет собой достаточно редкое случайное явление при нахождении процесса в статистически управляемом состоянии.

Контрольные карты Шухарта бывают двух основных типов: для количественных и альтернативных данных. Для каждой контрольной карты встречаются две ситуации:

- стандартные значения не заданы;
- стандартные значения заданы.

Стандартные значения - значения, установленные в соответствии с некоторыми конкретными требованиями или целями.

Контрольные карты для количественных данных:

1) карты среднего ($\overline{X}$) и размахов (R) или выборочных стандартных отклонений (s);

2) карта индивидуальных значений (X) и скользящих размахов (R);

3) карта медиан (Me) и размахов (R).

Контрольные карты для альтернативных данных:

1) карта долей несоответствующих единиц продукции (p) или карта числа несоответствующих единиц (np);

2) карта числа несоответствий (c) или карта числа несоответствий, приходящихся на единицу продукции (u).

Контрольные карты для количественных данных

Количественные данные представляют собой наблюдения, полученные с помощью измерения и записи значений некоторой характеристики для каждой единицы, рассматриваемой в подгруппе, например длина в метрах, сопротивление в омах, шум в децибелах и т.д. Карты для количественных данных, и особенно простейшие из них ($\overline{X}$- и R-карты), - это классические контрольные карты, применяемые для управления процессами.

Для контрольных карт, использующих количественные данные, предполагается нормальное (гауссово) распределение для вариаций внутри выборок, причем отклонения от этого предположения влияют на эффективность карт. Коэффициенты для вычисления контрольных границ выведены при условии нормальности. Поскольку контрольные границы используются только как эмпирические критерии при принятии решений, целесообразно пренебрегать малыми отклонениями от нормальности.

Благодаря центральной предельной теореме выборочные средние имеют распределение, приближающееся к нормальному с ростом объема выборки. Это обосновывает возможность предположения о нормальности для $\overline{X}$-карт даже при объемах выборок, столь малых как 4 или 5 единиц, взятых для проведения контроля. Если используют отдельные наблюдения для изучения возможностей процесса, истинное распределение важно. Рекомендуется периодически перепроверять выполнение таких предположений, чтобы убедиться, что используемые данные принадлежат одной совокупности.

Карты средних ($\overline{X}$) и размахов (R) или выборочных стандартных отклонений (s)

Карты для количественных данных отражают состояние процесса через разброс (изменчивость от единицы к единице) и через расположение центра (среднее процесса). Поэтому контрольные карты для количественных данных почти всегда применяют и анализируют парами - одна карта для расположения и одна - для разброса. Наиболее часто используют пару $\overline{X}$- и R-карту. В таблице 10 приведены формулы контрольных границ. Коэффициенты для соответствующих карт приведены в Приложении 3.

Формулы контрольных границ для карт Шухарта с использованием количественных данных

Таблица 10

Статистика	Стандартные значения не заданы		Стандартные значения заданы	
	Центральная линия	UCL и LCL	Центральная линия	UCL и LCL
$\overline{X}$	$\overline{\overline{X}}$	$\overline{\overline{X}} \pm A_2\overline{R}$ или $\overline{\overline{X}} \pm A_3\bar{s}$	X_0 или μ	$X_0 \pm A\sigma_0$
R	$\overline{R}$	$D_3\overline{R}$, $D_4\overline{R}$	R_0 или $d_2\sigma_0$	$D_1\sigma_0$, $D_2\sigma_0$
s	$\bar{s}$	$B_4\bar{s}$, $B_3\bar{s}$	s_0 или $C_4\sigma_0$	$B_5\sigma_0$, $B_6\sigma_0$
Примечание - Заданы стандартные значения X_0 или μ , R_0 , S_0 или σ_0 .				

Пример:

Пусть производятся контрольные измерения посадочных диаметров деталей класса валов, изготавливаемых на вашем предприятии (стандартные значения не заданы). В таблице 11 представлен результат измерений посадочных диаметров вала Ø$35h7^{0}_{-0,025}$.

Измерения производятся ежесменно.

Таблица 11

k				
1	**2**	**3**	**4**	**5**
Xik				
34,992	34,995	34,991	34,985	34,994
34,989	34,984	34,982	34,993	34,991
34,986	34,986	34,991	34,985	34,987
34,994	34,995	34,982	34,980	34,987
34,985	34,993	34,982	34,984	34,978
34,993	34,980	34,981	34,995	34,983
34,990	34,992	34,987	34,993	34,994
34,988	34,989	34,993	34,982	34,986
34,988	34,989	34,991	34,991	34,985
34,990	34,988	34,981	34,991	34,983
34,986	34,985	34,988	34,987	34,985
34,984	34,984	34,987	34,986	34,983
34,990	34,990	34,984	34,988	34,986
34,989	34,991	34,989	34,988	34,986
34,988	34,988	34,993	34,985	34,992
34,983	34,989	34,986	34,993	34,986
34,989	34,992	34,987	34,982	34,986
34,981	34,995	34,983	34,987	34,992
34,990	34,988	34,987	34,987	34,981
34,985	34,986	34,991	34,988	34,986
34,978	34,985	34,991	34,998	34,985
34,992	34,979	34,977	34,993	34,990
34,985	34,979	34,989	34,985	34,985
34,981	34,990	34,990	34,989	34,984
34,990	34,983	34,982	34,979	34,982

Для каждого *i*-го цикла измерений рассчитывается среднее значение результатов измерений $\bar{X}_i$ и размах результатов измерений R_i.(Таблица 11).

Таблица 11

Номер цикла измерений i	**Среднее значение**	**Наибольшее значение в цикле**	**Наименьшее значение в цикле**	**Размах**
1	34,991	34,995	34,985	0,009
2	34,988	34,993	34,982	0,012
3	34,987	34,991	34,985	0,006
4	34,988	34,995	34,980	0,015
5	34,984	34,993	34,978	0,014
6	34,986	34,995	34,980	0,016
7	34,991	34,994	34,987	0,007
8	34,987	34,993	34,982	0,011
9	34,989	34,991	34,985	0,006
10	34,987	34,991	34,981	0,009
11	34,986	34,988	34,985	0,003
12	34,985	34,987	34,983	0,004
13	34,988	34,990	34,984	0,006
14	34,988	34,991	34,986	0,005
15	34,989	34,993	34,985	0,009
16	34,988	34,993	34,983	0,010
17	34,987	34,992	34,982	0,010
18	34,988	34,995	34,981	0,013
19	34,987	34,990	34,981	0,009
20	34,987	34,991	34,985	0,006
21	34,987	34,998	34,978	0,020
22	34,986	34,993	34,977	0,016
23	34,985	34,989	34,979	0,009
24	34,987	34,990	34,981	0,009
25	34,983	34,990	34,979	0,012

Среднее результатов всех измерений $\bar{\bar{X}}$ =34,987 мм

Средний размах $\bar{R}$ =0,00098

Контрольные границы для построения $\bar{X}$ -карты

По таблице Приложения3 находим коэффициент $A_2 = 0{,}577$

Верхняя контрольная граница $UCL_{\bar{X}} = \bar{\bar{X}} + A_2 \cdot \bar{R}$ =34,993

Нижняя контрольная граница $LCL_{\bar{X}} = \bar{\bar{X}} - A_2 \cdot \bar{R}$ =34,981

Контрольная карта средних значений представлена на рисунке 43.

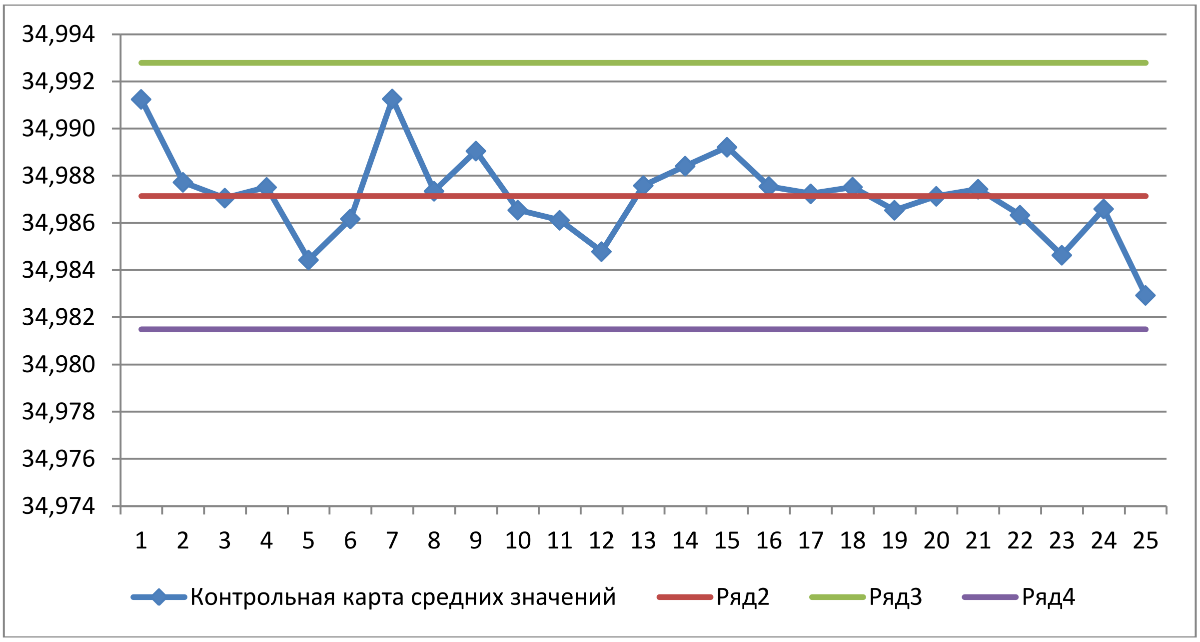


Рис. 43

Контрольные границы для построения R - карты

По таблице Приложения 3 находим коэффициент $D_4 = 2{,}114$

По таблице Приложения 3 находим коэффициент $D_3 = 0$

Верхняя контрольная граница $UCL_R = D_4 \cdot \bar{R}$ =0,0207

Нижняя контрольная граница $LCL_R = D_3 \cdot \bar{R}$ =0

Контрольная карта размахов представлена на рисунке 44.

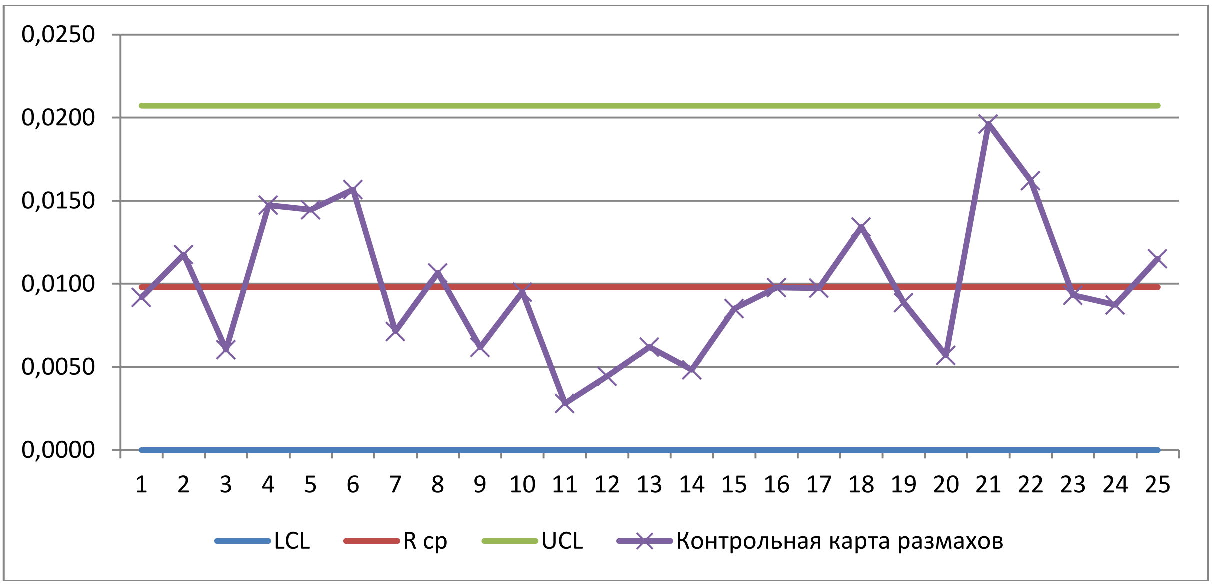


Рис. 44

Для построения s- карты в каждом цикле измерений рассчитывается среднее квадратическое отклонение (Таблица 12).

$$s = \sqrt{\frac{\sum_{k=1}^{5}(X_k - \overline{X})^2}{n-1}} \tag{55}$$

Таблица 12

Номер цикла измерений i	**s**
1	0,0015
2	0,0020
3	0,0010
4	0,0028
5	0,0022
6	0,0029
7	0,0011
8	0,0016
9	0,0010
10	0,0017
11	0,0005
12	0,0007
13	0,0010
14	0,0007
15	0,0014
16	0,0015
17	0,0015
18	0,0023
19	0,0014
20	0,0009
21	0,0030
22	0,0031
23	0,0014
24	0,0016
25	0,0017

Центральная линия $\bar{s}$ =0,0016

По таблице Приложения 3 находим коэффициент $B_3 = 0$

По таблице Приложения 3 находим коэффициент $B_4 = 2{,}089$

Верхняя контрольная граница $UCL_s = B_4\bar{s}$ =0,017,

Нижняя контрольная граница $LCL_s = B_3\bar{s}$ =0

s-карта представлена на рисунке 45.

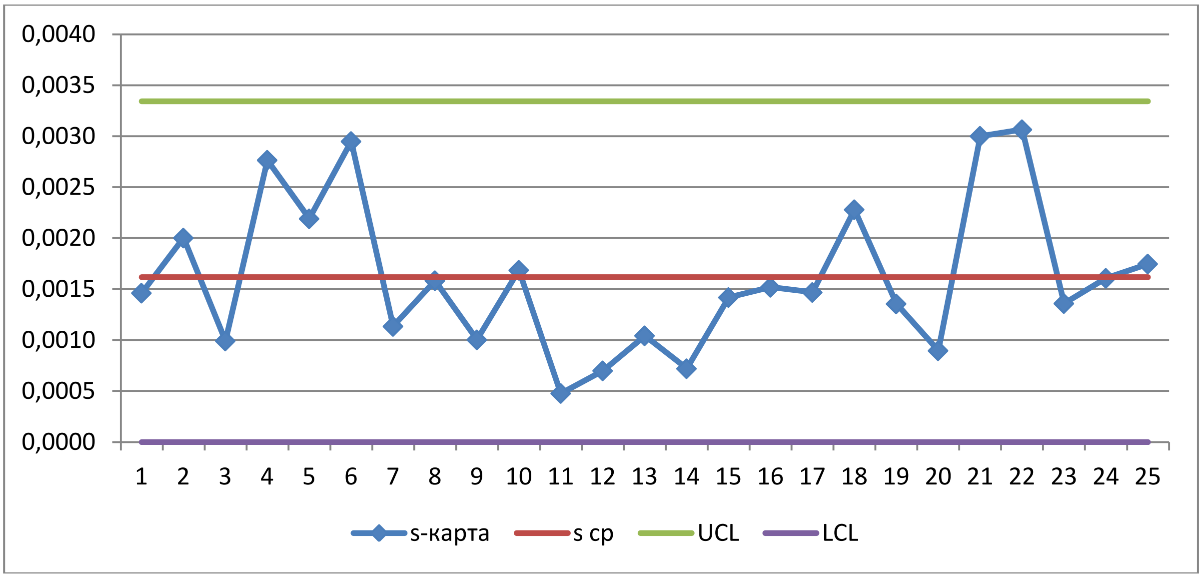


Рис. 45

Критерии оценки хода процесса

Для интерпретации хода процесса по картам Шухарта существует набор из восьми дополнительных критериев, который схематически показан на рисунках 46-53

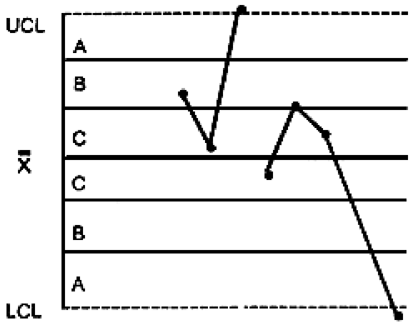


Рис. 46

КРИТЕРИЙ 1 - Одна точка вне зоны A

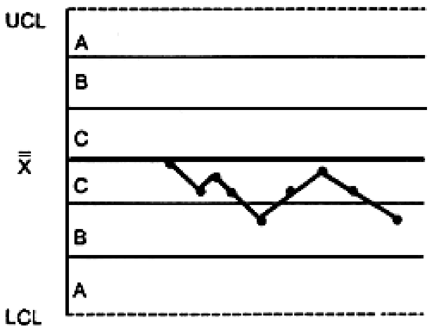


Рис.47

КРИТЕРИЙ 2 - Девять точек подряд в зоне C или по одну сторону от центральной линии

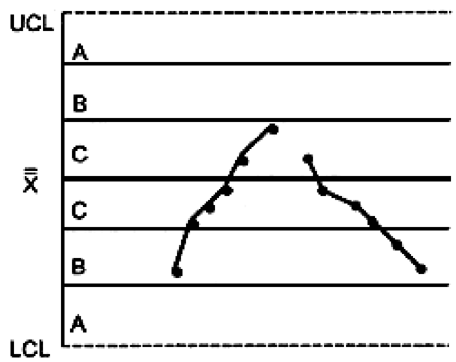


Рис. 48

КРИТЕРИЙ 3 - Шесть возрастающих или убывающих точек подряд

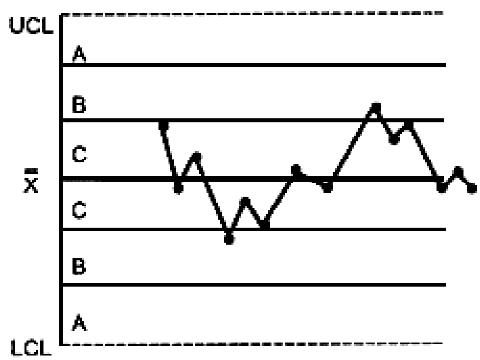


Рис. 49

КРИТЕРИЙ 4 - Четырнадцать попеременно возрастающих и убывающих точек

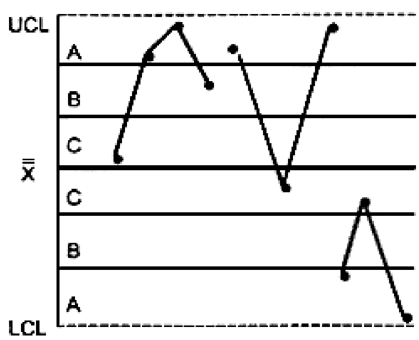


Рис. 50

КРИТЕРИЙ 5 - Две из трех последовательных точек в зоне A или вне ее

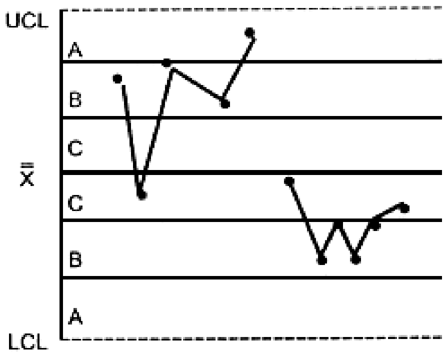


Рис. 51

КРИТЕРИЙ 6 - Четыре из пяти последовательных точек в зоне B или вне ее

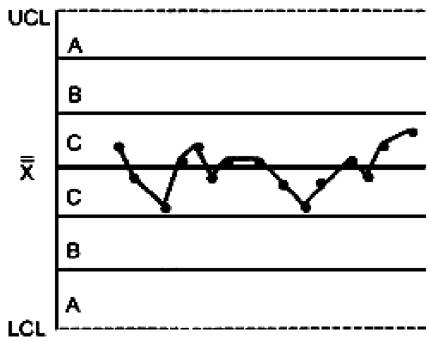


Рис. 52

КРИТЕРИЙ 7 - Пятнадцать последовательных точек в зоне C выше и ниже центральной линии

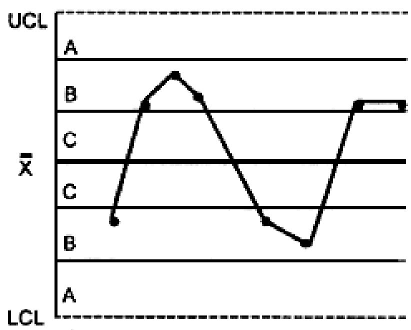


Рис. 53

КРИТЕРИЙ 8 - Восемь последовательных точек по обеим сторонам центральной линии и ни одной в зоне C.

Этот набор критериев можно принять за основу, но пользователи должны обращать внимание на любую необычную структуру точек, которая может указывать на проявление особых (неслучайных) причин. Поэтому эти критерии следует рассматривать только как примеры ситуаций, когда может быть установлено проявление неслучайных причин. Появление любого из случаев, описанных в этих критериях, - указание на присутствие особых причин, которые должны быть проанализированы и скорректированы.

Верхняя и нижняя контрольные границы установлены на расстоянии 3σ над и под центральной линией. Для применения этих критериев контрольная карта делится на шесть равных зон шириной σ. Эти зоны обозначаются A, B, C, C, B, A, причем зоны C расположены симметрично центральной линии. Данные критерии применимы к $\bar{x}$ - картам и X-картам индивидуальных значений. Предполагается нормальное распределение соответственно $\bar{X}$ и индивидуальных значений.

Оценка возможностей процесса

В стандарте ГОСТ Р 50779.42-99 рассмотрена методика оценки возможностей процесса путем определения индекса пригодности

$$\mathrm{PCI} = \frac{\text{допуск}}{\text{разброс процесса}} \frac{\mathrm{UTL} - \mathrm{LTL}}{6\hat{\sigma}}, \qquad (56)$$

где UTL - верхнее предельно допустимое значение контролируемого параметра;

LTL - нижнее предельно допустимое значение контролируемого параметра;

- оценивают по средней изменчивости внутри подгрупп и выражают как $\bar{S}/c_4$ или $\bar{R}/d_2$ (таблица Приложения 3).

При PCI меньше 1 возможности процесса неприемлемы, а при PCI, равном 1, процесс находится на грани требуемых возможностей. На практике в качестве минимально приемлемого значения берется PCI = 1,33, поскольку всегда есть некоторые вариации в выборках, и нет процессов, которые всегда находятся в статистически управляемом состоянии.

Рассчитайте индекс возможностей процесса изготовления валов с посадочным диаметром $\varnothing 35h7^{0}_{-0,025}$

$$UTL = 35$$

$$LTL = 34{,}975$$

$$\hat{\sigma} = \frac{\bar{s}}{C_4} = 0{,}0016 * 1{,}0638 = 0{,}0017$$

$$PCI(C_p) = \frac{35 - 34{,}975}{6 * 0{,}0017} = 2{,}42$$

Следует отметить, что PCI измеряет только отношение разброса процесса и допуска, а положение или центрирование процесса не учитывается. При высоких значениях PCI все-таки возможен выход доли значений за установленные пределы. Поэтому важно также оценивать расстояние между средним процесса и ближайшим предельно допустимым значением.

<u>Задание</u>

В MS Office Excel произведите расчеты и построить контрольные $\bar{X}$-карты, R-карты и s-карты выборки нормального распределения 25 циклов результатов контрольных измерений посадочных диаметров деталей класса валов, изготавливаемых на вашем предприятии (стандартные значения не заданы). Произведите анализ и оценку возможностей процесса.

Построение карты средних ($\bar{X}$) и размахов (R) в Statistica 6.0

Данные выборки примера построения карты средних ($\bar{X}$) и размахов (R) - результата измерений посадочных диаметров вала $\varnothing 35h7^{0}_{-0,025}$, сгенерированные в MS Excel расположите в один столбец, добавьте столбец с номером образца и столбец с номером группы. Сохраните файл в формате Excel 97-2003. Откройте сохраненный файл в программе Statistica 6.0 (рис. 54).

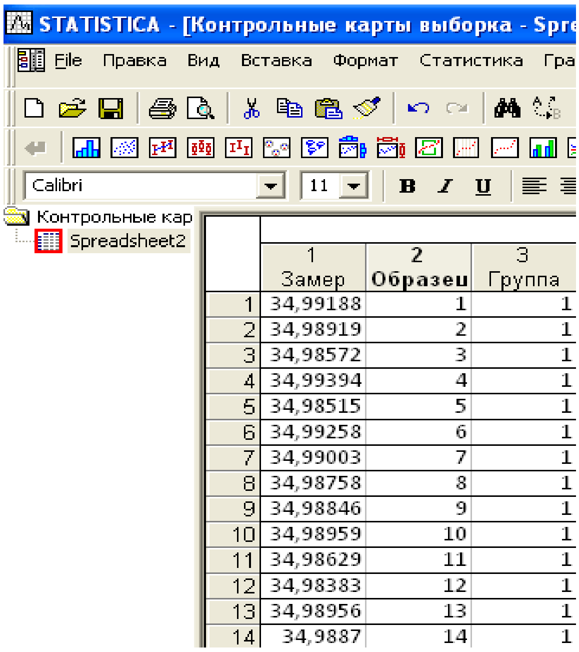


Рис. 54

Выберите в меню **Статистика-Индустриальная статистика&Сигма шесть-Качество диаграммы управления.** В открывшемся списке выберите **X-bar&R chat for variables**. Нажмите **ОК** (рис. 55).

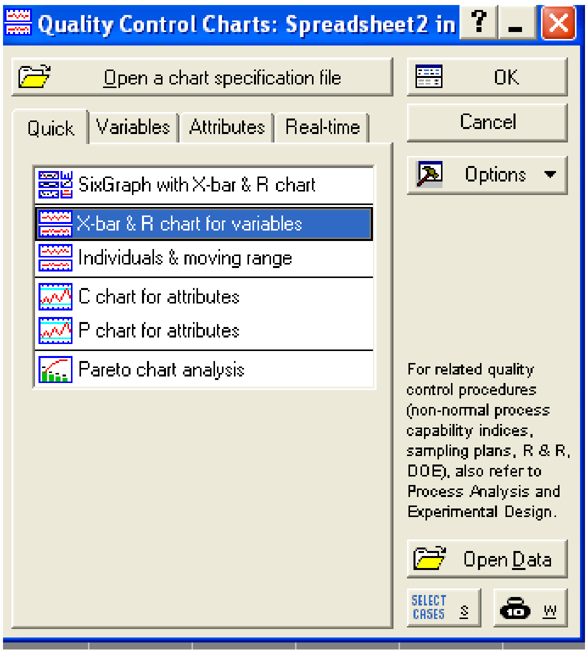


Рис. 55

В открывшемся диалоговом окне выберите переменные для анализа. Нажмите **ОК** (рис. 56).

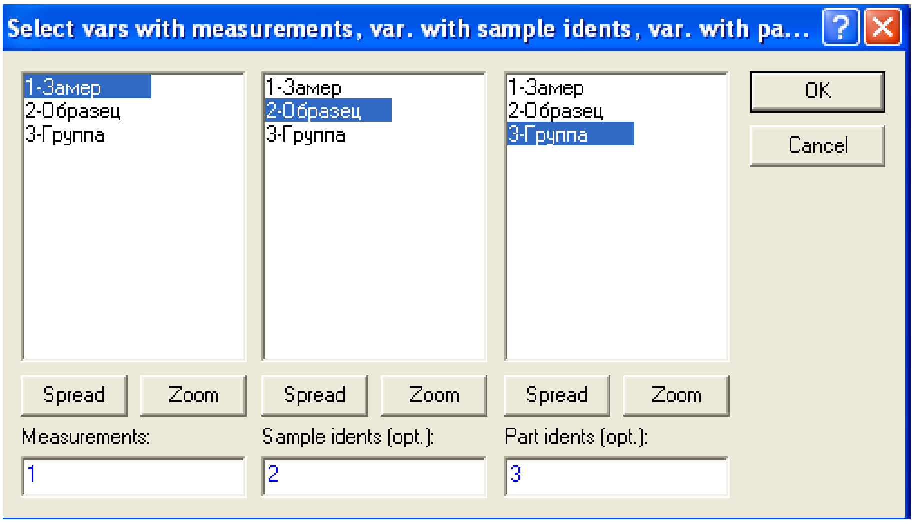


Рис. 56

Задайте количество образцов в группе. Нажмите **ОК** (рис. 57).

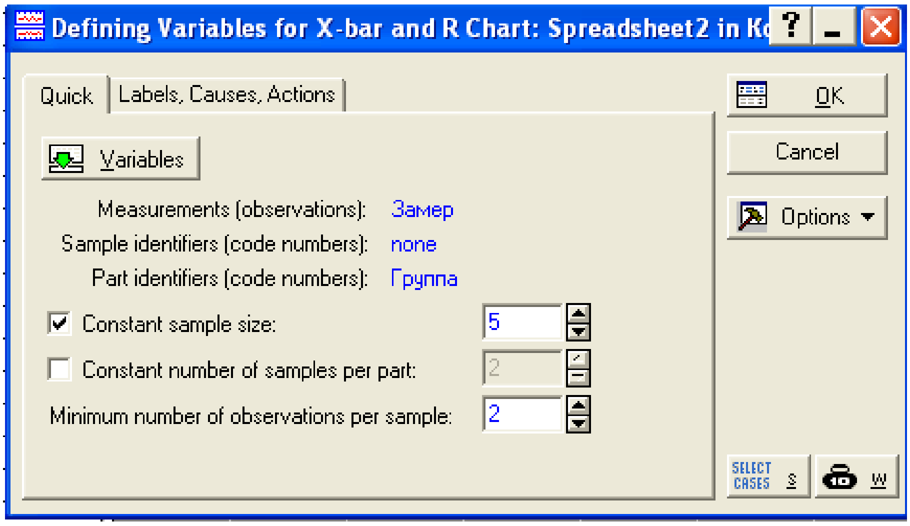


Рис. 57

Как видно на рисунке 58, помимо карты средних ($\overline{X}$) и размахов (R) построены гистограммы распределения этих параметров.

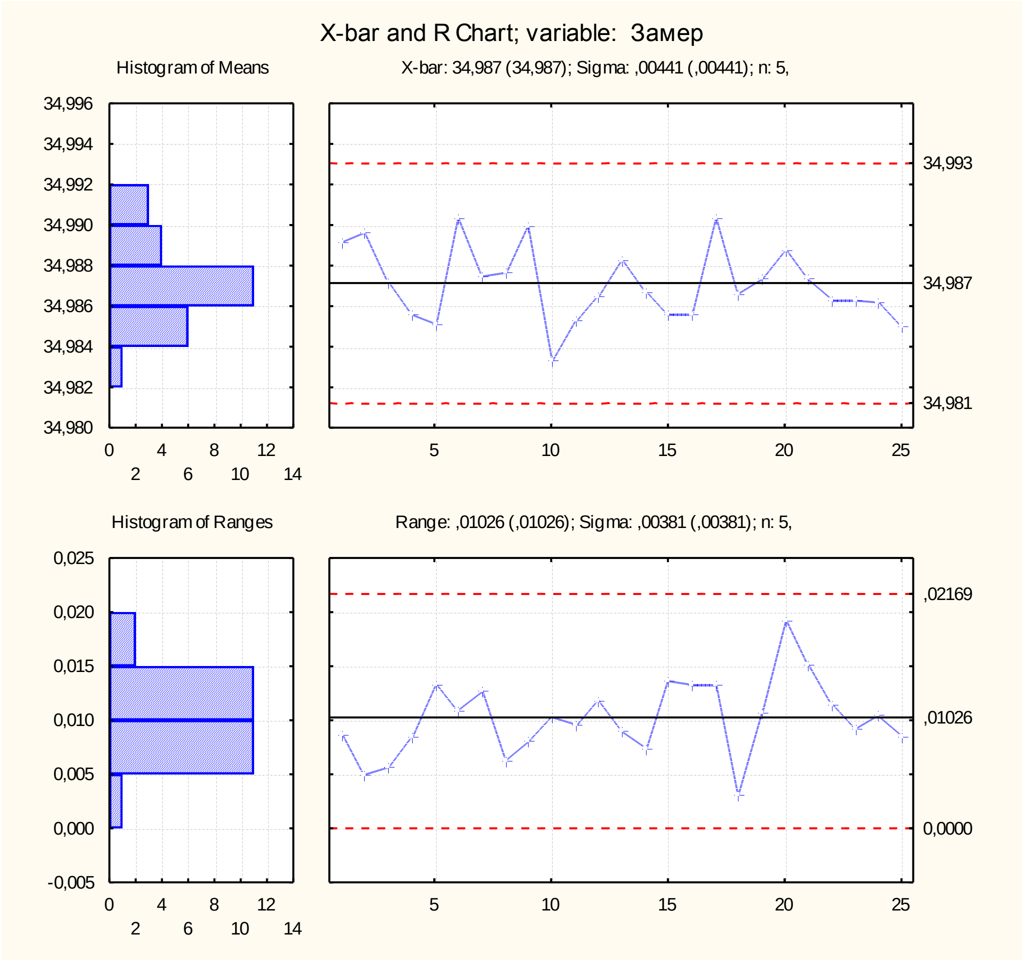


Рис. 58

Помимо этих карт программа формирует другие представления. Для этого необходимо нажать на соответствующие кнопки вкладки Charts (рис. 59).

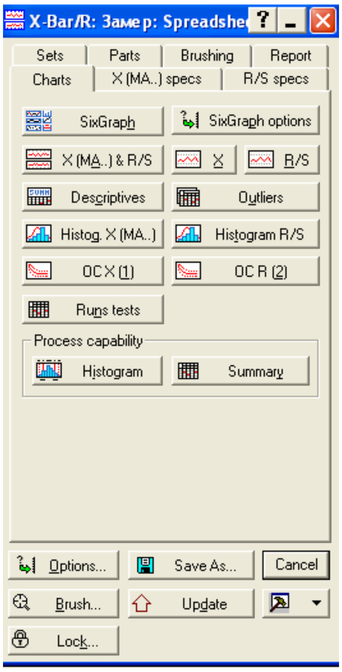


Рис. 59

При нажатии на кнопку **Outliers** программа формирует отчет о выходе значений контролируемого параметра за пределы контрольных границ (рис. 60).

Type of outlier	Number of outliers: S Замер; R-Chart; Avera Standard; Range: 0,0	
	Number of outlier	% of all samples
< LCL	0	0,00
> UCL	0	0,00
Total	0	0,00

Рис. 60

При нажатии на кнопку **RunsTests**программа оценку по принятым критериям (рис.61).

Zones A/B/C: 3,000/2,000/1,000 * Sigma Tests for special causes (runs rules)	Замер ; Runs Te R Chart Center line: 0,010	
	from sample	to sample
9 samples on same side of center	OK	OK
6 samples in row in/decreasing	OK	OK
14 samples alternating up & down	OK	OK
2 of 3 samples in Zone A or beyond	OK	OK
4 of 5 samples in Zone B or beyond	OK	OK
15 samples in Zone C	OK	OK
8 samples beyond Zone C	OK	OK

Рис. 61

Задание

В программе Statistica 6.0 постройте контрольные $\overline{X}$-карты и R-карты выборки нормального распределения 25 циклов результатов контрольных измерений посадочных диаметров деталей класса валов(стандартные значения не заданы) (Приложение 2). Произведите анализ и оценку возможностей процесса.

Определение предполагаемого истинного значения измеряемого параметра образца

1) Определение предполагаемого истинного значения измеряемого параметра X_0 осуществляется в метрологическом зале с использованием средства измерительной техники наиболее высокой точности.

2) В случае если выполнение условий п. 1 невозможно, рекомендуется выбрать из производства образец, значение измеряемого параметра которого попадает в середину интервала допуска, измерить этот образец 20 раз и за предполагаемое истинное значение взять среднее значение полученных измерений в условиях стабильности измерительного процесса (Таблица 13).

Таблица 13

Определение стандартного значения	
Номер измерения образца	Результат лабораторных измерений образца для вычисления истинного значения параметра
1	34,984
2	34,989
3	34,981
4	34,981
5	34,993
6	34,986
7	34,994
8	34,984
9	34,991
10	34,988
11	34,990
12	34,988
13	34,999
14	34,989
15	34,984
16	34,987
17	34,992
18	34,991
19	34,988
20	34,984
X_0	34,988

Построение $\overline{X}$ - карты, R-карты, - карты для известных стандартных значений

Пусть заданы стандартные значения параметров:

X_0 =34,988 мм

$\sigma_0 = 0{,}004$ мм

Контрольные границы для построения $\overline{X}$ - карты.

По таблице Приложения 3 находим коэффициент $A_1 = 1{,}342$.

Верхняя контрольная граница $UCL = X_0 + A_1\sigma_0$ =34,993.

Нижняя контрольная граница $LCL = X_0 - A_1\sigma_0$ =34,983.

$\overline{X}$ - карта для известных стандартных значений представлена на рисунке 62.

Рис. 62

Контрольные границы для построения R - карты

По таблице Приложения 3 находим коэффициент $d_2 = 2{,}326$

По таблице Приложения 3 находим коэффициент $D_1 = 0$

По таблице Приложения 3 находим коэффициент $D_2 = 4{,}918$

Центральная линия $d_2\sigma_0$=0,0093

Верхняя контрольная граница $UCL_r = D_2\sigma_0$=0,020,

Нижняя контрольная граница $LCL_r = D_1\sigma_0 = 0$

R - карта для известных стандартных значений представлена на рисунке 63.

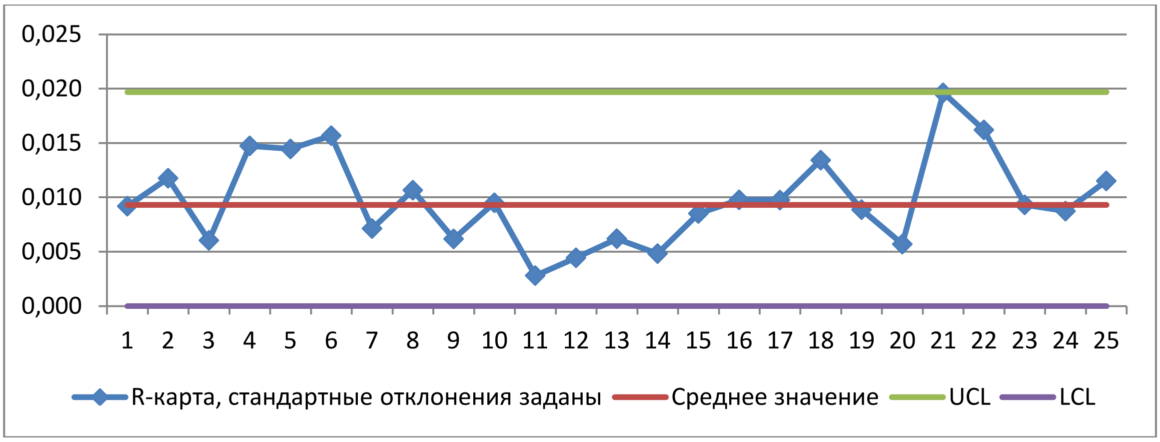


Рис. 63

Контрольные границы для построения S - карты.

По таблице Приложения 3 находим коэффициент C_4=0,94.

По таблице Приложения 3 находим коэффициент B_5 =0.

По таблице Приложения 3 находим коэффициент B_6 =1,964.

Центральная линия $C_4\sigma_0$=0,00376.

Верхняя контрольная граница $B_6\sigma_0$ =0,0079

Нижняя контрольная граница $B_5\sigma_0$ =0

s - карта для известных стандартных значений представлена на рисунке 64.

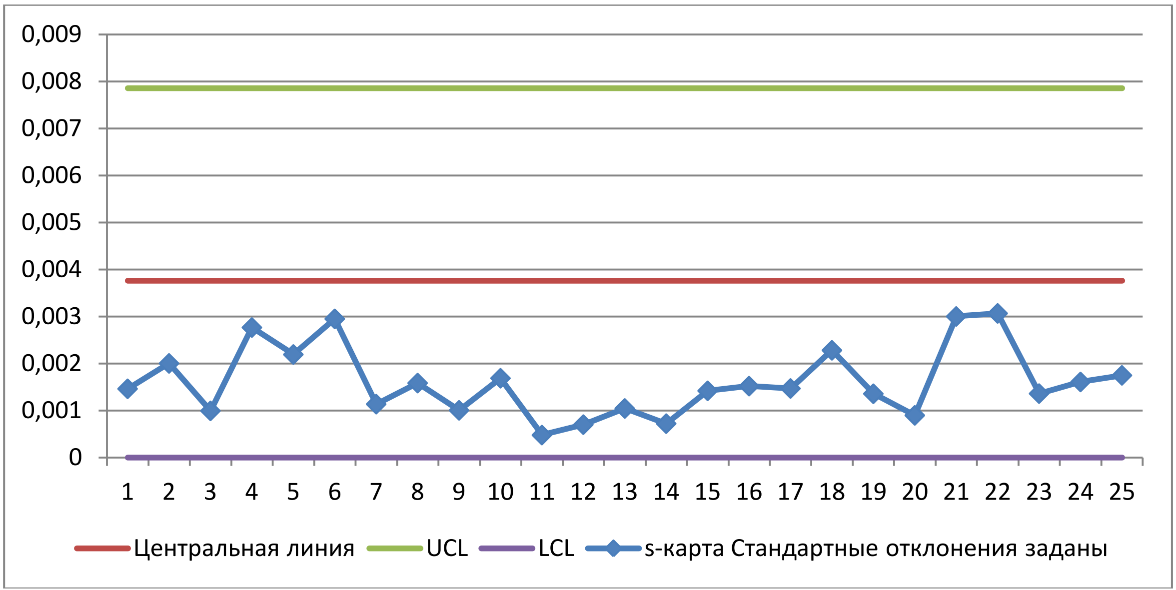


Рис. 64

<u>Задание</u>

В MS Office Excelпостроить контрольные $\overline{X}$ -карты, R-карты и s-карты для выборки нормального распределения 25 циклов результатов контрольных измерений посадочных диаметров деталей класса валов (Приложение 2). В случае, если точки на картах вышли за контрольные границы, их необходимо исключить из дальнейших расчетов и для стабильного измерительного процесса (оставшиеся значения) рассчитать $\overline{\overline{X}}$ и $\overline{s}$. Принять $X_0 = \overline{\overline{X}}$ мм, $\sigma_0 = \overline{s}$, мм и построить $\overline{X}$ -карты, R-карты и s-карты для известных значений X_0 и σ_0.

Контрольные карты индивидуальных значений

В некоторых ситуациях для управления процессами невозможно либо непрактично иметь дело с рациональными подгруппами. Время, или стоимость, требуемые для измерения при одиночном наблюдении, столь велики, что проведение повторных наблюдений даже не рассматривают. Это обычно происходит, когда измерения дорогостоящие (например при разрушающем контроле) или выход продукции все время относительно однороден. В других ситуациях нельзя получить более одного значения, например показание прибора или значение характеристики партии исходных материалов, поэтому приходится управлять процессом на основе индивидуальных значений.

При использовании карт индивидуальных значений рациональные подгруппы для обеспечения оценки изменчивости внутри партии не применяют и контрольные границы рассчитывают на основе меры вариации, полученной по скользящим размахам обычно двух наблюдений. Скользящий размах - это абсолютное значение разности измерений в последовательных парах, т.е. разность первого и второго измерений, затем второго и третьего и т.д.

На основе скользящих размахов вычисляют средний скользящий размах $\overline{R}$, который используют для построения контрольных карт. Также по всем данным вычисляют общее среднее $\overline{\overline{X}}$. В таблице 14 приведены формулы расчета контрольных границ для карт индивидуальных значений.

Формулы контрольных границ для карт индивидуальных значений

Таблица 14

Статистика	Стандартные значения не заданы		Стандартные значения заданы	
	Центральная линия	UCL и LCL	Центральная линия	UCL и LCL
Индивидуальное значение X	$\overline{\overline{X}}$	$\overline{\overline{X}} \pm E_2\overline{R}$	X_0 или μ	$X_0 \pm 3\sigma_0$
Скользящий размах R	$\overline{R}$	$D_4\overline{R}$, $D_3\overline{R}$	R_0 или $d_2\sigma_0$	$D_2\sigma_0$, $D_1\sigma_0$

Примечания

1 Заданы стандартные значения X_0 и R_0 или μ и σ_0.

2 $\overline{R}$ обозначает среднее скользящего размаха из двух наблюдений (n = 2).

3 Значения коэффициентов d_2, D_1, D_2, D_3, D_4 и косвенно $E_2 = 3/d_2$ можно получить из таблицы Приложения 3 при n = 2.

При использовании карт индивидуальных значений необходимо учитывать следующее:

– карты индивидуальных значений не столь чувствительны к изменениям процесса, как $\overline{X}$ - и R-карты;

– при интерпретации карт индивидуальных значений следует проявлять осторожность, если распределение процесса не является нормальным;

– карты индивидуальных значений не оценивают повторяемость процесса от изделия к изделию, и поэтому в некоторых случаях лучше использовать обычные $\overline{X}$ - и R-карты с малыми объемами выборочных подгрупп (от 2 до 4), даже если это приведет к увеличению интервала между подгруппами.

Пример.

В таблице 15 представлены данные микробиологического лабораторного контроля показателя общей бактериальной обсемененности (КМАФАнМ) 10 партий молока, поступающего на молочный завод.

Таблица 15

Номер партии	**КМАФАнМ КОЕ/см3 молоко**
1	460000
2	490000
3	230000
4	250000
5	360000
6	320000
7	340000
8	410000
9	290000
10	490000

Построение карты индивидуальных значений X и скользящих размахов R Стандартные значения не заданы (Таблица 16).

Таблица 16

Номер партии	X	R
1	460000	-
2	490000	30000
3	230000	260000
4	250000	20000
5	360000	110000
6	320000	40000
7	340000	20000
8	410000	70000
9	290000	120000
10	490000	200000

$$\bar{\bar{X}} = 364000 \text{ КОЕ/см}^3$$

$$\bar{R} = 96667 \text{ КОЕ/см}^3$$

$$UCL = \bar{\bar{X}} + E_2\bar{R},$$

где $E_2 = 3/d_2$.

По таблице Приложения 3 находим коэффициент d_2=1,128 для n = 2

$$UCL = 621092$$

$$LCL = \bar{\bar{X}} - E_2\bar{R}$$

$$LCL = 106908$$

Карта индивидуальных значений X представлена на рисунке 65.

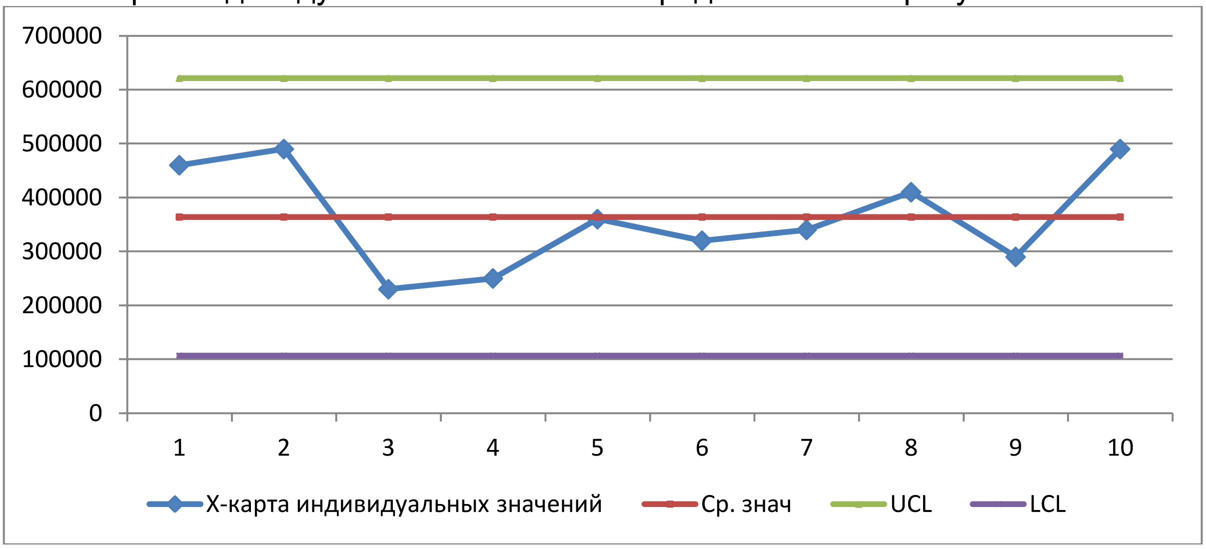


Рис. 65

Контрольные границы для построения карты скользящих размахов R

По таблице Приложения 3 находим коэффициент s D_4=3,267 и D_3=0,000 для n = 2

$$UCL = D_4\bar{R} = 315810$$

$$LCL = D_3\bar{R} = 0$$

Карта скользящих размахов R представлена на рисунке 66.

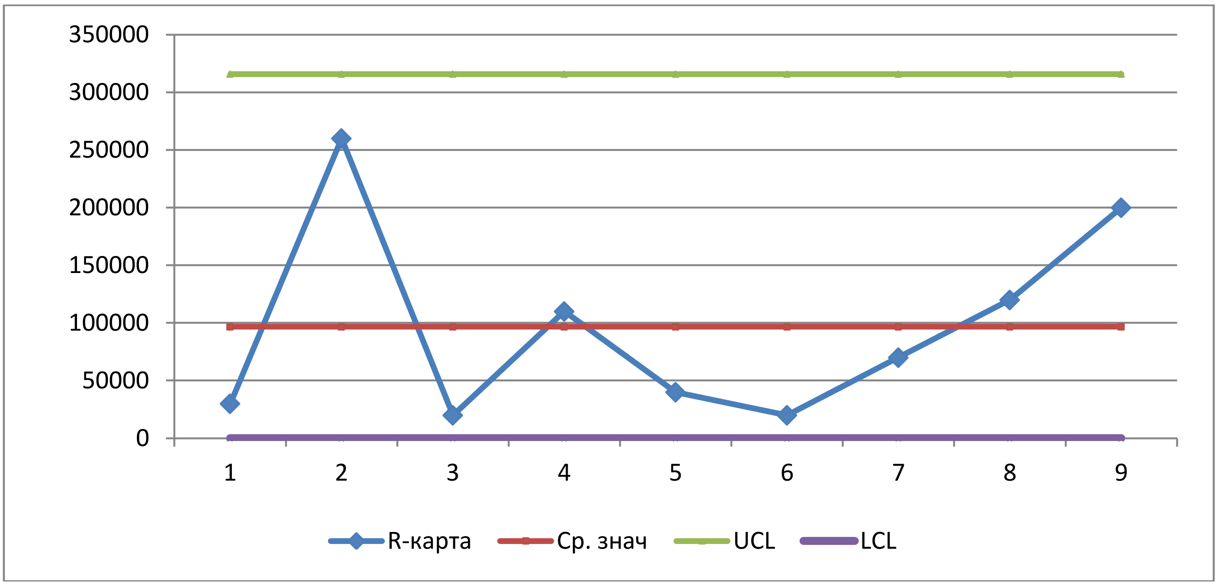


Рис. 66

> ***<u>Задание</u>***
> ***В MS Office Excel для выборки нормального распределения индивидуальных значений - результатов контроля 50 значений размера деталей построить карты индивидуальных значений X и скользящих размахов R.***

Построение карты индивидуальных значений X и скользящих размахов R в Statistica 6.0

В программе Statistica 6.0 откройте данные микробиологического лабораторного контроля показателя общей бактериальной обсемененности (КМАФАнМ) 10 партий молока, поступающего на молочный завод (рис. 67).

	1 Var1
1	460000
2	490000
3	230000
4	250000
5	360000
6	320000
7	340000
8	410000
9	290000
10	490000

Рис. 67

Выберите в меню **Статистика-Индустриальная статистика&Сигма шесть-Качество диаграммы управления.**

В открывшемся списке выберите **Individuals&moving range**. Нажмите **ОК**.

В открывшемся диалоговом окне выберите переменную Var1 для анализа. Нажмите **ОК** (рис. 68).

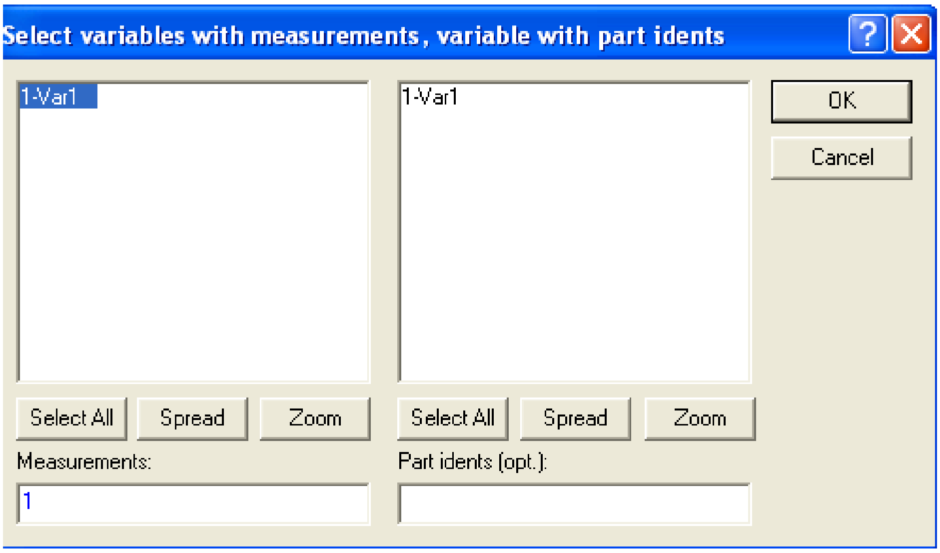


Рис. 68

Как видно на рисунке 69, помимо карты индивидуальных значений (X) и карты скользящих размахов ($\bar{R}$) построены гистограммы распределения этих параметров.

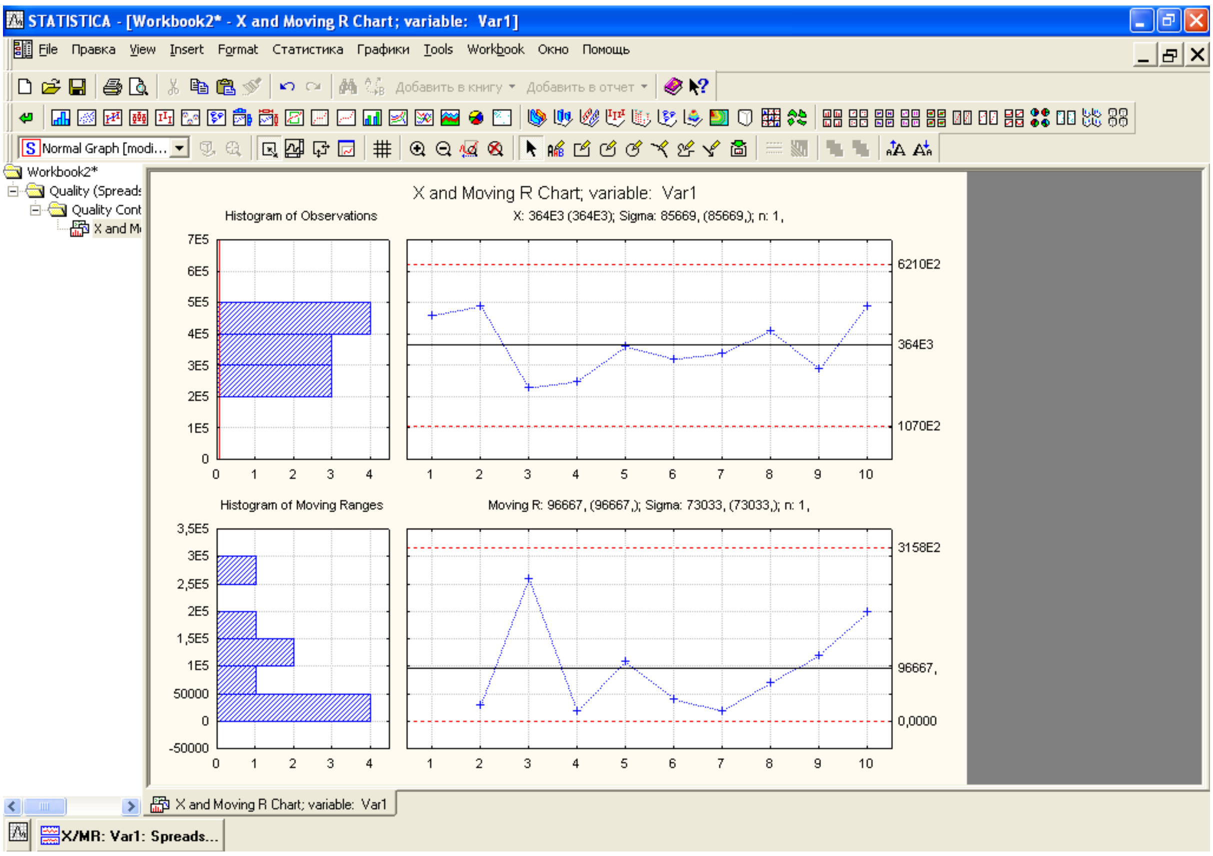


Рис. 69

Контрольные карты для альтернативных данных

Альтернативные данные представляют собой наблюдения, фиксирующие наличие или отсутствие некоторых характеристик (или признаков) у каждой единицы рассматриваемой подгруппы. На основе этих данных производится подсчет числа единиц, обладающих или не обладающих данным признаком, или число таких событий в единице продукции, группе или области. Альтернативные данные в общем случае могут быть получены быстро и дешево, для сбора их не требуется специального обучения.

Биномиальный закон распределения

МОДЕЛЬ 1. Рассмотрим модель барабана с черными и белыми шарами. Пусть p - вероятность того, что извлечен черный шар. Так как модель состоит из черных и белых шаров, то вероятность извлечь белый шар равна $1-p$. Каждый раз вынутый шар возвращают в барабан. Для нас неважно, появится ли сначала черный или белый шар, так как при анализе выборки порядок результатов не имеет значения.

Так как вероятность появления черного шара равна p, а белого $1-p=q$, вероятности приведенных результатов будут следующие:

$p^2, 2pq, q^2$

Вероятность появления первой, второй или третьей комбинации равна сумме отдельных вероятностей. Эта сумма вероятностей будет равна 1, так как перечисленные результаты составляют полную группу событий.

Вероятности перечисленных результатов можно получить путем разложения бинома $(p+q)^2$; поэтому такое распределение называют биномиальным.

Теоретически случаи появления черного, или белого шара могут быть представлены в виде распределения частот, как это показано на рисунке70.

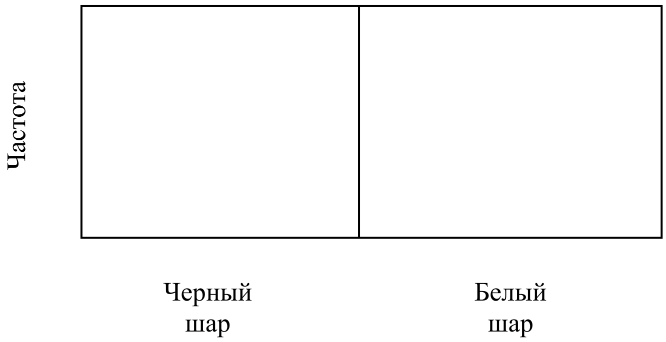


Рис. 70

При достаточно большом количестве извлечений шаров частота появления белого шара должна равняться частоте появления черного шара. Площадь, ограниченная соответствующей кривой распределения делится так, что 50% приходится на белые, а 50% на черные шары, а суммарная вероятность равна 1.

МОДЕЛЬ 2. Игральная кость имеет шесть граней, на каждую из которых нанесено определенное количество очков – от одного до шести. На рисунке показано теоретическое распределение частот выпадения граней игральной кости при достаточно большом количестве испытаний (рис. 71).

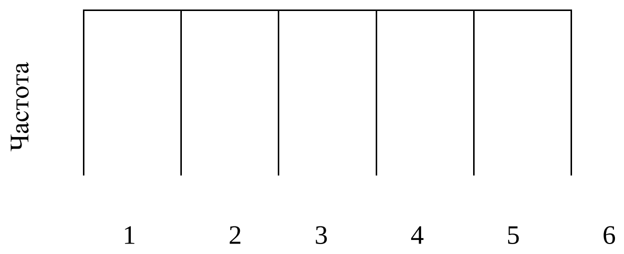


Рис. 71

Рисунок говорит о том, что вероятность появления каждой из шести граней равна 1/6, а суммарная вероятность равна 1.

МОДЕЛЬ 3. Пусть производится бросание пары игральных костей. На рисунке представлено теоретическое распределение относительной частоты появления суммарного числа очков, выраженное в %. Такое распределение обусловлено числом возможных сочетаний (рис. 72).

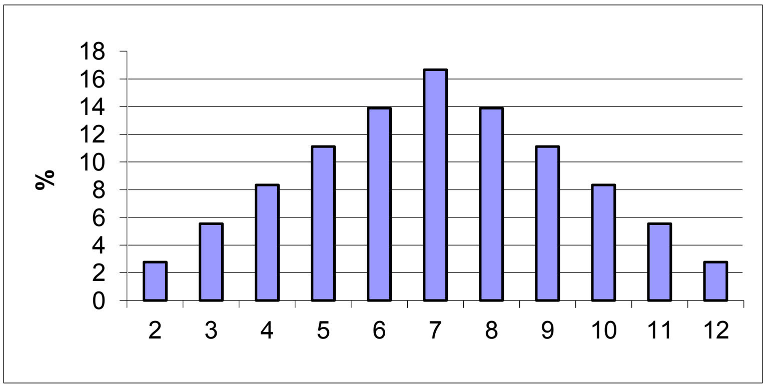


Рис. 72

Это распределение, как видно, существенно отличается от распределений, изображенных на предыдущих рисунках. Площади, расположенные ближе к краям, существенно меньше по своим размерам, чем площади, расположенные ближе к центру. На рисунке мы обнаруживаем так называемые хвосты распределения. Площадь, соответствующая появлению 12 очков, составляет 1/6 площади, соответствующей появлению 7 очков.

Для трех игральных костей распределение суммарного числа очков площади хвостов распределения еще более уменьшаются в размерах по отношению к площадям, расположенным ближе к центру. Например, вероятность появления 18 очков составляет только 1/27 вероятности появления 10 очков и 1/216 суммарной вероятности появления всех очков.

Каждое из приведенных распределений является дискретным (прерывным). Это означает, что между ординатами этих распределений не существует непрерывного плавного перехода, тогда как теоретическая кривая нормального распределения представляет собой непрерывную кривую. Тем не менее, это не связано ни с какими практическими трудностями, поскольку существует возможность приближенной нормализации такого рода распределений.

МОДЕЛЬ 4. Возвращаясь к примеру с черными и белыми шарами, допустим, что вместо одного шара производится одновременное извлечение 9 шаров (n) и притом большое количество раз. Очевидно, существует несколько возможных сочетаний белых и черных шаров, которые могут появиться в результате эксперимента, причем некоторые из этих сочетаний будут появляться чаще, чем другие. Число сочетаний белых и черных шаров может быть исчислено по формуле бинома Ньютона, где p' - вероятность появления белого шара, а q'-вероятность появления черного шара.

На рисунке дано графическое изображение теоретического биномиального распределения результатов такого испытания. Интересно отметить, что хотя значение величины $(p^{/})$ теоретически принимается равным 0,5, в действительности его величина обладает симметричным рассеянием.

Для применения нормального закона распределения необходимо располагать фактическими данными или оценками среднего значения и среднего квадратичного отклонения. Истинное среднее значение $(p^{/})$ в данном случае известно; оно равно 0,5 (рис. 73).

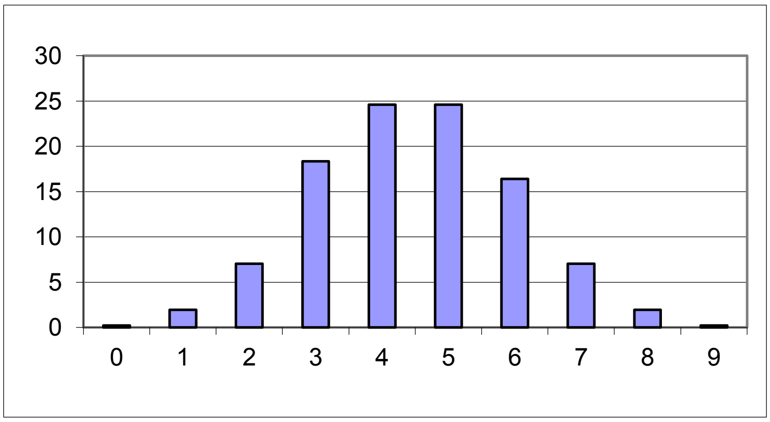


Рис. 73

Если «с возвращением» извлекать из барабана n шаров, то вероятность возможного значения $X = k$ (числа появлений события - появления белого шара) рассчитывается по формуле Бернулли:

$$P_n(k) = C_n{}^k p^k q^{n-k}, \tag{57}$$

где $C_n{}^k$ - число сочетаний из n элементов по k, равное $\dfrac{n!}{k!(n-k)!}$.

Математическое ожидание числа появления белого шара биномиального распределения равно произведению числа испытаний на вероятность появления события в каждом испытании: $M(X) = np$.

Среднее квадратичное отклонение (именуемое также стандартным отклонением) биномиального распределения может быть выражено:

- через число событий

$$\sigma_{np} = \sqrt{np'(1-p')} \tag{58}$$

- в виде доли событий

$$\sigma_p = \frac{\sqrt{p'(1-p')}}{\sqrt{n}} \tag{59}$$

- в виде процентного отношения событий

$$\sigma_{100p} = \frac{\sqrt{100\,p'(100-100\,p')}}{\sqrt{n}} \tag{60}$$

Под словом «событие» здесь подразумевается наименование конкретного события – в данном случае – появление белого шара.

На практике подобной моделью пользуются для определения числа бракованных деталей в выборке, подразумевая под событием появление дефектного изделия.

Параметр q заменяется здесь на $1-p$. Поскольку $p+q=1$, то

$q=1-p$ и $p+(1-p)=1$.

Характеристики биномиального распределения p' и σ_p аналогичны характеристикам $\overline{X}'$ и $\sigma_{\overline{X}}$ нормального распределения.

Для нашего примера $\sigma_p = \frac{\sqrt{0{,}50 \times 0{,}50}}{\sqrt{9}} = 0{,}167$.

Кривая нормального распределения очень близко приближается к биномиальной кривой при $p'=0{,}5$ и $n \ge 10$.

Контрольная карта долей несоответствующих единиц (p-карта)

p-Карта измеряет долю несоответствующих единиц в контролируемой группе. Это может быть выборка определенного объема, отбираемая дважды в день; процент продукции, группируемой на почасовой или ежедневной основе; доля поставок точно в срок и т.д. Можно контролировать один или несколько показателей качества.

До применения *p*-карты необходимо сделать следующее:

1) создать среду, подходящую для работы;
2) определить процесс. Процесс должен быть понятен и взаимоувязан с другими операциями,
3) определить характеристики, подлежащие управлению, выделить из них наиболее перспективные для совершенствования процесса (применение диаграммы Парето). При этом рекомендуется проанализировать:

- потребности потребителей последующего процесса производства и потребителей конечной единицы продукции;
- области текущих и потенциальных проблем. Рассматривают имеющиеся факты потерь или низкой эффективности (например брак, переделки, затраты времени, недостижимые цели) и области риска (например многократные изменения продукции или услуг и элементов процесса);
- корреляцию между характеристиками. Для эффективного и результативного изучения процесса используют преимущества взаимосвязи между характеристиками. Если несколько отдельных характеристик изделия имеют тенденцию изменяться вместе, может быть целесообразно строить карту

только для одной из них;

4) определить измерительную систему. Характеристики должны быть однозначно определены, так чтобы данные могли быть доступны всем заинтересованным лицам (какая информация, где, когда и при каких условиях собрана). Определение характеристик может повлиять на тип используемых контрольных карт;

5) минимизировать изменчивость от особых причин. Цель этого – избежать очевидных проблем, которые могут и должны быть решены без контрольных карт. Во всех случаях необходимо вести записи, отмечающие все существенные события, такие как процедурные изменения, новый исходный материал и т.п. Это поможет в последующем анализе процесса.

При использовании контрольных карт для альтернативных данных достаточно одной карты, так как предполагаемое распределение имеет только один независимый параметр - средний уровень. p - и np -карты основаны на биномиальном распределении, а c - и u -карты - на распределении Пуассона.

Расчеты для этих карт одинаковы, за исключением случаев непостоянства объема подгрупп. Когда объем подгрупп постоянен, для каждой подгруппы могут быть выбраны одни и те же контрольные границы. Если число контролируемых единиц в каждой подгруппе различно, должны быть рассчитаны контрольные границы отдельно для каждого объема подгруппы. Таким образом, np - и c -карты могут быть применены при постоянном объеме подгруппы, а p - и u -карты - в любой ситуации.

Когда объем подгруппы изменяется от выборки к выборке, для каждой подгруппы рассчитывают свои контрольные границы, при этом, чем меньше объем подгруппы, тем шире полоса между этими границами, и наоборот. Если объем подгрупп меняется несущественно, то можно ограничиться одним набором контрольных границ, основанным на среднем объеме подгруппы. Для практических целей достаточно, если объемы подгрупп находятся в пределах ±25% целевого объема подгруппы.

Обычно p -карту используют для определения среднего процента несоответствующих единиц, обнаруженных за определенный период времени. Она привлекает внимание персонала процесса и управляющих к любым изменениям этого среднего. Процесс признается находящимся в состоянии статистической управляемости также, как и при использовании $\overline{X}$ - и R -карт. Если все выборочные точки ложатся внутри пробных контрольных границ без выбросов, указывающих на наличие особых причин, то о процессе можно сделать заключение, что он управляем. В этом случае средняя доля несоответствующих $\overline{p}$ единиц берется как стандартное значение для доли несоответствующих единиц p_0 (Таблица 17).

Формулы контрольных границ карт Шухарта для альтернативных данных

Таблица 17

Статистика	Стандартные значения не заданы		Стандартные значения заданы	
	Центральная линия	3σ -е контрольные границы	Центральная линия	3σ -е контрольные границы
p	$\bar{p}$	$\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$	p_0	$p_0 \pm 3\sqrt{\frac{p_0(1-p_0)}{n}}$
np	$n\bar{p}$	$n\bar{p} \pm 3\sqrt{n\bar{p}(1-\bar{p})}$	np_0	$np_0 \pm 3\sqrt{np_0(1-p_0)}$
c	$\bar{c}$	$\bar{c} \pm 3\sqrt{\bar{c}}$	c_0	$c_0 \pm 3\sqrt{c_0}$
u	$\bar{u}$	$\bar{u} \pm 3\sqrt{\frac{\bar{u}}{n}}$	u_0	$u \pm 3\sqrt{\frac{u_0}{n}}$
Примечание - p_0, np_0, c_0, u_0 - заданные стандартные значения.				

Когда $\bar{p}$ мало и (или) n мало, значение LCL_p – отрицательно. В этих случаях нижней контрольной границы нет, поскольку даже значение p, равное 0 для конкретного периода, находится внутри границ случайной изменчивости.

Пример построения p - Карты (стандартные значения не заданы):

В таблице 18 указано число несоответствующих единиц, найденных при сплошном контроле изделий, изготавливаемых на автоматизированной линии. Поскольку неисправность серьезна, для определения момента выхода линии из статистически управляемого состояния используют контрольную карту процента несоответствующих единиц - карта получена при сборе предварительных данных по 25 подгруппам каждая из 800 изделий.

Таблица 18

Номер подгруппы	**Число несоответствующих деталей**	**Число проконтролированных деталей**	**Процент несоответствий**
1	33	800	4,13%
2	34	800	4,25%
3	31	800	3,88%
4	31	800	3,88%
5	23	800	2,88%
6	24	800	3,00%
7	27	800	3,38%
8	33	800	4,13%
9	37	800	4,63%
10	29	800	3,63%
11	32	800	4,00%
12	23	800	2,88%
13	21	800	2,63%
14	26	800	3,25%
15	32	800	4,00%
16	35	800	4,38%
17	27	800	3,38%
18	31	800	3,88%
19	36	800	4,50%
20	25	800	3,13%
21	24	800	3,00%
22	23	800	2,88%
23	37	800	4,63%
24	24	800	3,00%
25	25	800	3,13%
ИТОГО	723	20000	

Центральная линия и контрольные границы:

Центральная линия: $\bar{p}=\frac{723}{2000} = 0{,}03615 \cong 3{,}62\%$

$$\mathrm{UCL} = \bar{p} + 3\sqrt{\bar{p}(1-\bar{p})/n}$$

$$\mathrm{UCL} = 0{,}03615 + 3\sqrt{\frac{0{,}03615(1-0{,}03615)}{800}} = 0{,}05595 \cong 5{,}59\%$$

$$\mathrm{LCL} = \bar{p} - 3\sqrt{\bar{p}(1-\bar{p})/n}$$

$$UCL = 0{,}03615 - 3\sqrt{\frac{0{,}03615(1 - 0{,}03615)}{800}} = 0{,}01635 \cong 1{,}64\%$$

p -карта представлена на рисунке 74.

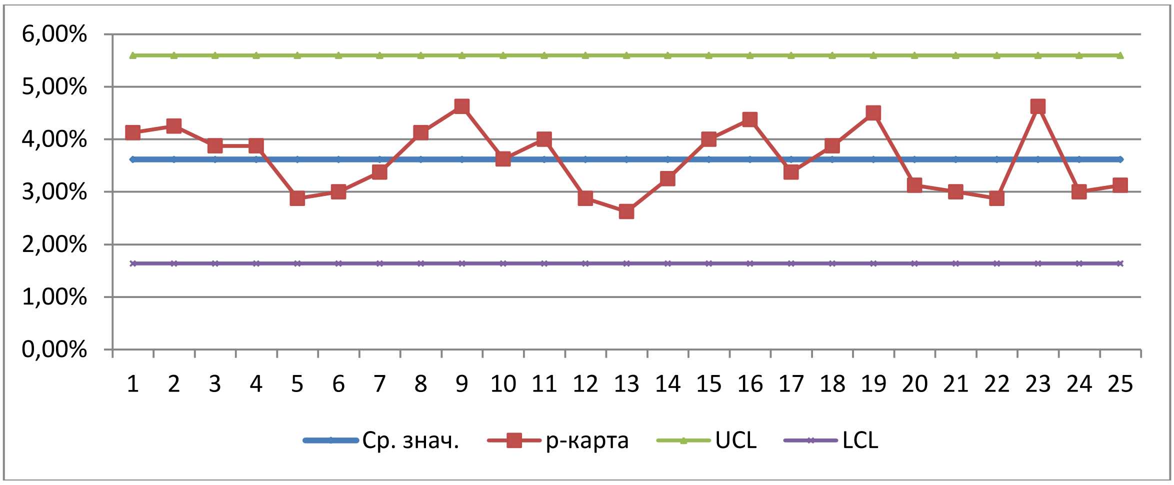


Рис. 74

Генерация выборки биномиального распределения в MS Office Excel

Генерация выборки биномиального распределения в **MS Office Excel** осуществляется в следующей последовательности.

Шаг 1. Зайдите в меню **Сервис - Анализ данных - Генерация случайных чисел.** Нажмите **ОК.**

Шаг 2. В появившемся окне выберите число переменных 1, число случайных чисел – 25, распределение выберите из раскрывающегося списка «Биномиальное» с параметрами значение р = 0,05, Число испытаний = 800и сгенерируйте выборку в выходном интервале A1. Нажмите **ОК.**(рис. 75).

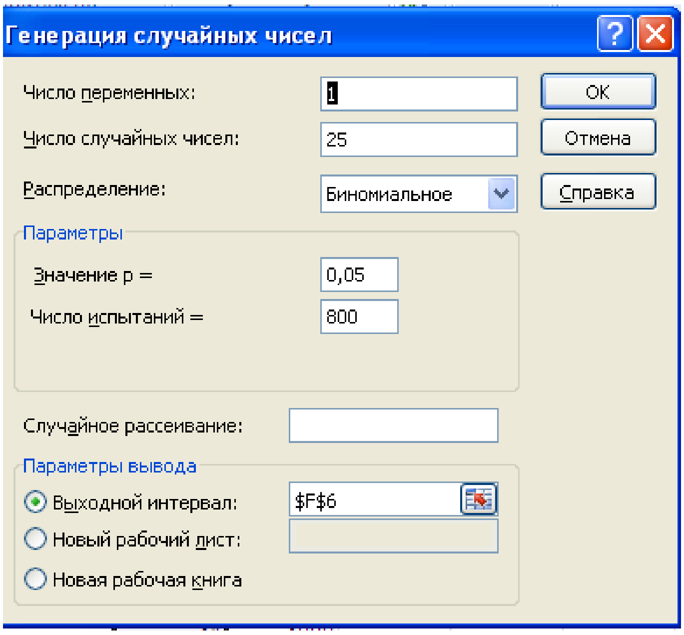


Рис. 75

В результате произведенных действий будет сгенерирован столбец данных – число несоответствующих деталей в 25 циклах контроля партий 800 деталей.

.

<u>Задание</u>
В MS Office Excel с помощью надстройки Анализ данных сформируйте выборку биномиального распределения числа несоответствующих деталей в 25 циклах контроля партий деталей. Значение вероятности брака – p и размер контролируемой партии указаны в варианте задания.

Построение p-карты в Statistica 6.0

Откройте данные контроля 25 подгрупп размером по 800 изделий в программе Statistica 6.0 (рис. 76).

	1 Var1	2 Var2	3 Var3
1	33	800	
2	34	800	
3	31	800	
4	31	800	
5	23	800	
6	24	800	
7	27	800	
8	33	800	
9	37	800	
10	29	800	
11	32	800	
12	23	800	
13	21	800	
14	26	800	
15	32	800	
16	35	800	
17	27	800	
18	31	800	
19	36	800	
20	25	800	

Рис. 76

Выберите в меню **Статистика-Индустриальная статистика&Сигма шесть-Качество диаграммы управления.**

В открывшемся списке выберите **P chart for attributes**. Нажмите **OK**.

В открывшемся диалоговом окне выберите переменную Var1 для анализа и Var2 - количество контролируемых изделий. Нажмите **OK** (рис. 77).

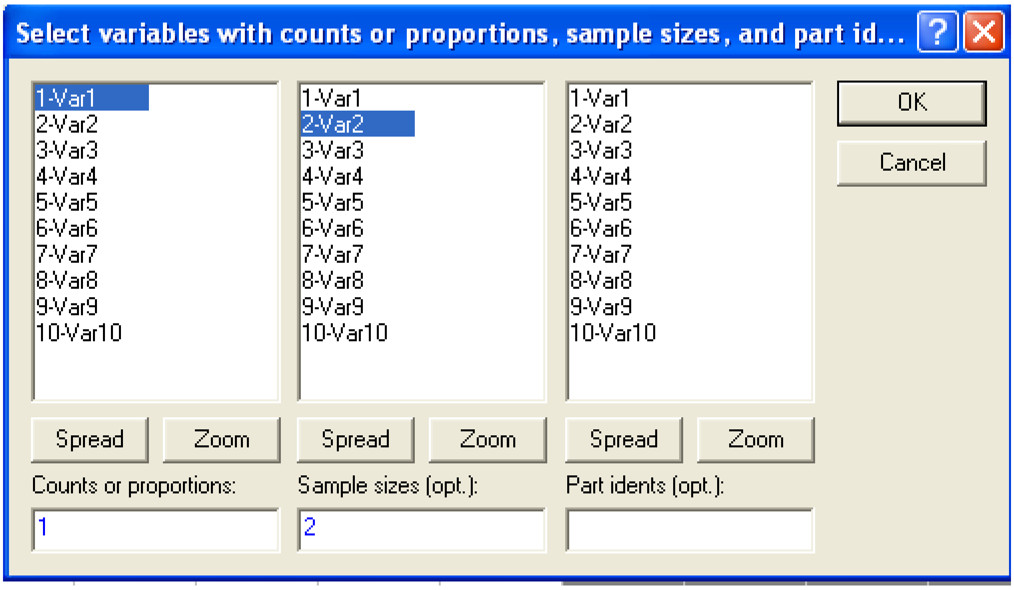


Рис. 77

Как видно на рисунке 78, помимо р-карты построена гистограмма распределения бракованных изделий.

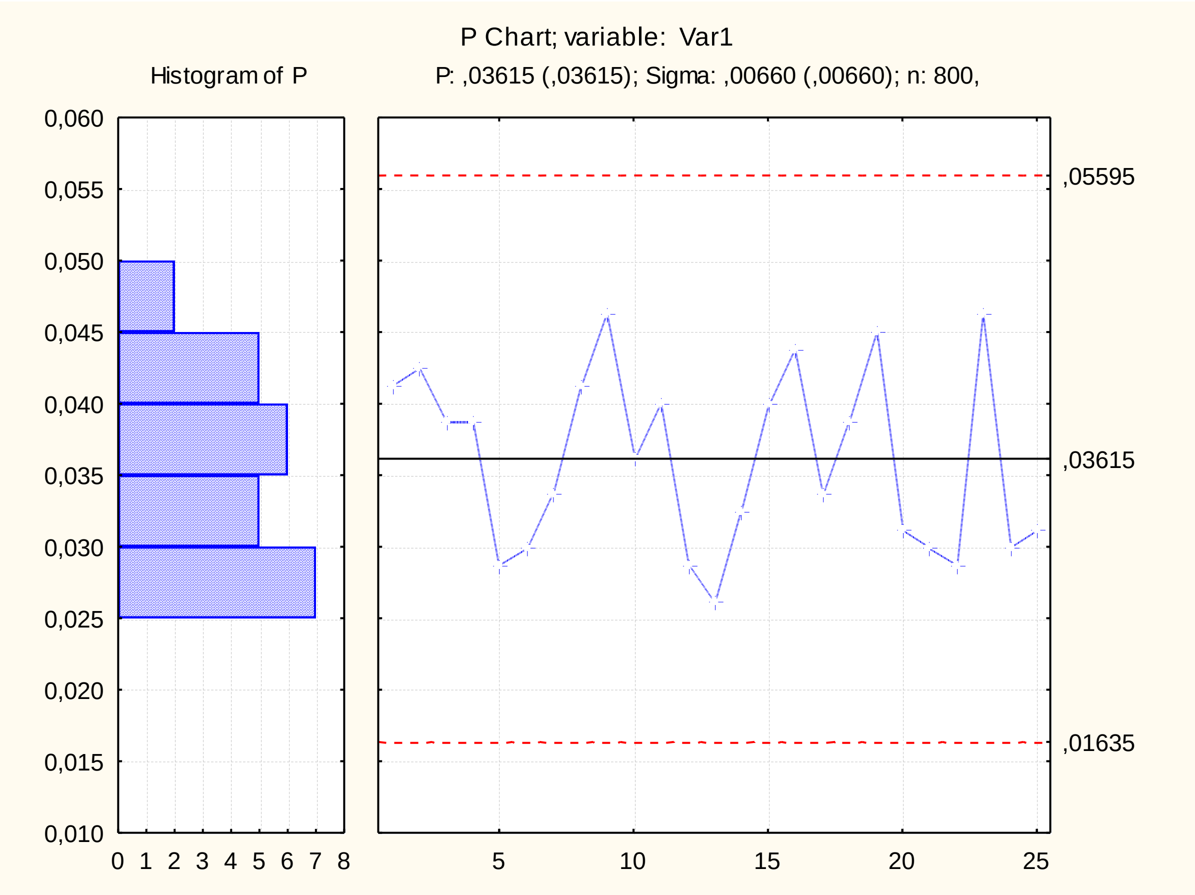


Рис. 78

КОНТРОЛЬНЫЕ ЛИСТКИ

Стабильный процесс - это процесс, в котором отсутствует системная вариация, то есть процесс, который последовательно производит продукцию, обладающую одинаковыми свойствами. Стабильный процесс имеет нормальное распределение частоты значений. Все это видно на построенной гистограмме. Стабильность означает заранее известные результаты, что гарантирует нам качество.

Для обеспечения стабильности процесса необходимо обеспечить контроль его параметров. Для сбора исходных данных используются контрольные листки.

Контрольный листок (или лист) — это инструмент для сбора данных и автоматического их упорядочения для облегчения дальнейшего использования собранной информации.

Обычно контрольный листок представляет собой бумажный бланк, на котором заранее напечатаны контролируемые параметры, согласно которым можно заносить в листок данные с помощью пометок или простых символов. Он позволяет автоматически упорядочить данные без их последующего переписывания. Таким образом, контрольный листок — хорошее средство регистрации данных.

Например, нам необходимо исследовать дефектные изделия в течение одного месяца. В данном контролируемом изделии встречается какое-то количество различных дефектов. Таким образом, наиболее целесообразным будет классификация данных по типам дефектов. Нечасто встречающиеся типы дефектов, следует объединить под общим заголовком “прочие”. Разрабатывается контрольный листок для регистрации данных с перечнем видов собираемой информации. В нем предусматривается место для графической регистрации данных.

Контрольные листки могут применяться как при контроле по качественным, так и при контроле по количественным признакам. Для каждой конкретной цели может быть разработан свой листок. Но принцип их оформления остается неизменным. При составлении контрольных листков следует обратить внимание на то, чтобы было указано, кто, на каком этапе процесса и в течение какого времен собирал данные, а также, чтобы форма листка была простой и понятной без дополнительных пояснений. Важно и то, чтобы все данные добросовестно фиксировались, и собранная в контрольном листке информация могла быть использована для анализа процесса (рис. 79).

Наименование Документа	Контрольный листок по видам дефектов	
Предприятие: XXX **Цех:** ________________ **Участок:** ____________	**Изделие:** ______________________ **Операция:** ____________________ **Контролер:** ___________________	**Кол-во Деталей** _____
Типы дефектов	***Данные контроля***	***ИТОГО***
Деформации	///// ///// ///// ///// ///// ///// ///// ///// ///// //	47
Царапины	///// ///// ///// ///// ///// ///// ///// ///// //	42
Трещины	///// ///// ///// ///// ////	24
Раковины	///// ///// ///// ///// ///// ///// ///// ///	38
Пятна	///// ///// ///// ///// ///// ///// ///// ///// ///// ///// ///	53
Разрыв	///// //	7
Прочие	///// ///// //	12
ИТОГО		

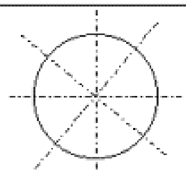
Наименование Документа	Контрольный листок по месту расположения дефектов	
Предприятие: XXX **Цех:** ________________ **Участок:** ____________	**Изделие:** ______________________ **Операция:** ____________________ **Контролер:** ___________________	**Кол-во Деталей** _____
		
Типы дефектов	***Данные контроля***	***ИТОГО***
Деформации	///// ///// ///// ///// ///// ///// ///// ///// ///// //	47
Царапины	///// ///// ///// ///// ///// ///// ///// ///// //	42
Трещины	///// ///// ///// ///// ////	24
Раковины	///// ///// ///// ///// ///// ///// ///// ///	38
Пятна	///// ///// ///// ///// ///// ///// ///// ///// ///// ///// ///	53
Разрыв	///// //	7
Прочие	///// ///// //	12
ИТОГО		

Рис. 79

На основании информации строятся диаграммы Парето. В таблице 19 представлены данные контрольного листка дефектов пайки, выявленные в течение месяца на участке изготовления печатных плат.

Таблица 19

Дефекты пайки	Число дефектов
Непропаянный шов	82
«Корявый» шов	10
Наплывы припоя	15
Излом в месте спая	10
Припой не смачивает поверхность паяемого металла	22
Припой при хорошей смачиваемости шва не затекает в зазор	45
Трещины в шве	10
Смещения и перекосы в паяных соединениях	45
Прочее	5
ИТОГО:	**244**

ДИАГРАММА ПАРЕТО

Проблемы качества оборачиваются потерями (дефектные изделия и затраты, связанные с их производством). Чрезвычайно важно прояснить картину распределения потерь. Большинство из них будет обусловлено незначительным числом видов дефектов, вызванных небольшим количеством причин. Таким образом, выяснив причины появления немногочисленных существенно важных дефектов, можно устранить почти все потери, сосредоточив усилия на ликвидации именно этих причин и отложив пока рассмотрение причин, приводящих к остальным многочисленным несущественным дефектам. Такого рода проблема успешно решается с помощью диаграммы Парето.

Диаграмма Парето — инструмент, позволяющий распределить усилия для разрешения возникающих проблем и выявить основные причины, с которых нужно начинать действовать.

В повседневной деятельности по контролю и управлению качеством постоянно возникают всевозможные проблемы, связанные, например, с появлением брака, неполадками оборудования, увеличением времени от выпуска партии изделий до ее сбыта, наличием на складе нереализованной продукции, поступлением рекламаций. Диаграмма Парето позволяет распределить усилия для разрешения возникающих проблем и установить основные факторы, с которых нужно начинать действовать с целью преодоления возникающих проблем.

Различают два вида диаграмм Парето:

1. Диаграмма Парето по результатам деятельности. Эта диаграмма предназначена для выявления главной проблемы и отражает следующие нежелательные результаты деятельности:

- качество: дефекты, поломки, ошибки, отказы, рекламации, ремонты, возвраты продукции;
- себестоимость: объем потерь, затраты;
- сроки поставок: нехватка запасов, ошибки в составлении счетов, срыв сроков поставок;
- безопасность: несчастные случаи, трагические ошибки, аварии.

2. Диаграмма Парето по причинам. Эта диаграмма отражает причины проблем, возникающих в ходе производства, и используется для выявления главной из них:

- исполнитель работы: смена, бригада, возраст, опыт работы, квалификация, индивидуальные характеристики;
- оборудование: станки, агрегаты, инструменты, оснастка, организация использования, модели, штампы;
- сырье: изготовитель, вид сырья, завод-поставщик, партия;
- метод работы: условия производства, заказы-наряды, приемы работы, последовательность операций;
- измерения: точность (указаний, чтения, приборная), верность и повторяемость (умение дать одинаковое указание в последующих измерениях одного и того же значения), стабильность (повторяемость в течение длительного периода), совместная точность, т.е. вместе с приборной точностью и тарированием прибора, тип измерительного прибора (аналоговый или цифровой).

Построение диаграммы Парето осуществляется в следующей последовательности.

1) Решите, какие проблемы надлежит исследовать и как собирать данные.

Какого типа проблемы вы хотите исследовать?

Пример: дефектные изделия, потери в деньгах, несчастные случаи.

Какие данные надо собрать и как их классифицировать?

Пример: по видам дефектов, по месту их появления, по процессам, по станкам, по рабочим, по технологическим причинам.

Примечание. Суммируйте остальные нечасто встречающиеся признаки под общим заголовком "прочие".

Установите метод и период сбора данных.

Примечание. Если это рекомендуется, используйте специальный бланк.

2) Разработайте контрольный листок для регистрации данных с перечнем видов собираемой информации. В нем надо предусмотреть место для графической регистрации данных проверок.

3) Заполните листок регистрации данных и подсчитайте итоги.

4) Для построения диаграммы Парето разработайте бланк таблицы для проверок данных, предусмотрев в нем графы для итогов по каждому проверяемому признаку в отдельности, накопленной суммы числа дефектов, процентов к общему итогу и накопленных процентов.

5) Расположите данные, полученные по каждому проверяемому признаку, в порядке значимости и заполните таблицу.

6) Примечание. Группу "прочие" надо поместить в последнюю строку вне зависимости от того, насколько большим получилось число, так как ее составляет совокупность признаков, числовой результат по каждому из которых меньше, чем самое маленькое значение, полученное для признака, выделенного в отдельную строку.

7) Начертите одну горизонтальную и две вертикальные оси.

1) Вертикальные оси:

(а) левая ось. Нанесите на эту ось шкалу с интервалами от 0 до числа, соответствующего общему итогу;

(б) правая ось. Нанесите на эту ось шкалу с интервалами от 0 до 100%.

2) Горизонтальная ось. Разделите эту ось на интервалы в соответствии с числом контролируемых признаков.

8) Постройте столбиковую диаграмму. Начертите кумулятивную кривую (кривую Парето). На вертикалях, соответствующих правым концам каждого интервала на горизонтальной оси, нанесите точки накопленных сумм (результатов или процентов) и соедините их между собой отрезками прямых.

9) Нанесите на диаграмму все обозначения и надписи:

– надписи, касающиеся диаграммы (название, разметка числовых значений на осях, наименование контролируемого изделия, имя составителя диаграммы);

– надписи, касающиеся данных (период сбора информации, объект исследования и место проведения, общее число объектов контроля).

Построение диаграммы Парето средствами MS Excel

Постройте таблицу (20) данных дефектов пайки. Для этого упорядочите по убыванию данные контрольного листка дефектов пайки. Добавьте столбцы, в которых произведите расчет накопленной суммы дефектов, процент числа дефектов по каждому признаку к общей сумме и накопленный процент дефектов.

Таблица 20

Номер	Дефекты пайки	Число дефектов	Накопленная сумма числа дефектов	Процент числа дефектов по каждому признаку к общей сумме	Накопленный процент
1	Непропаянный шов	82	82	33,61%	33,61%
2	Припой при хорошей смачиваемости шва не затекает в зазор	45	127	18,44%	52,05%
3	Смещения и перекосы в паяных соединениях	45	172	18,44%	70,49%
4	Припой не смачивает поверхность паяемого металла	22	194	9,02%	79,51%
5	Наплывы припоя	15	209	6,15%	85,66%
6	«Корявый» шов	10	219	4,10%	89,75%
7	Излом в месте спая	10	229	4,10%	93,85%
8	Трещины в шве	10	239	4,10%	97,95%
9	Прочие	5	244	2,05%	100,00%
	ИТОГО:	244		100,00%	

Постройте гистограмму числа дефектов пайки (рис. 80).

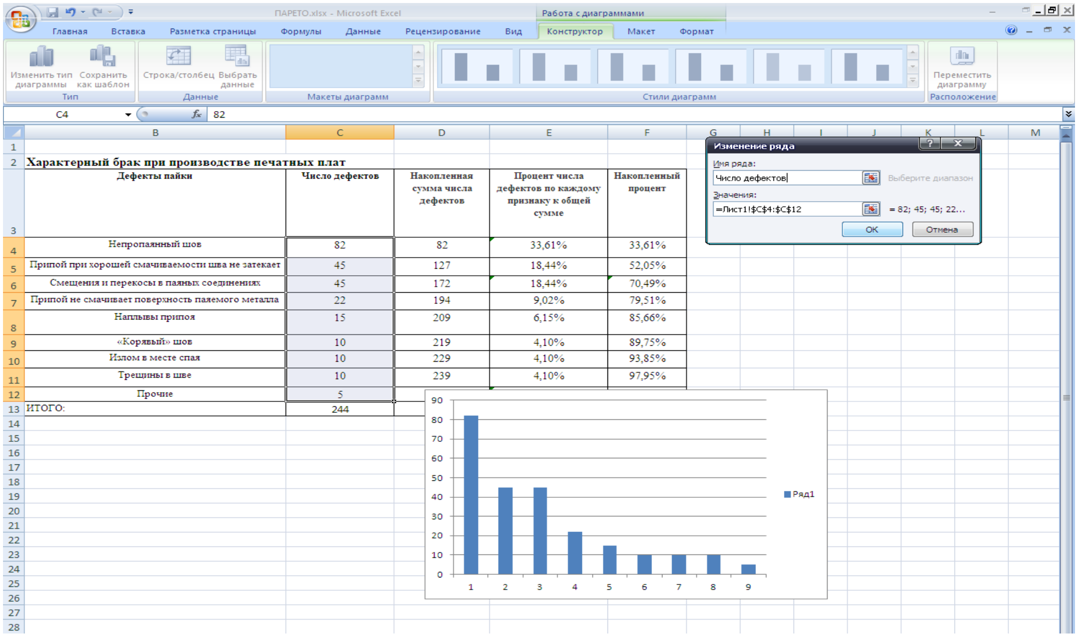


Рис. 80

Добавьте ряд данных «Накопленная сумма дефектов» (рис. 81).

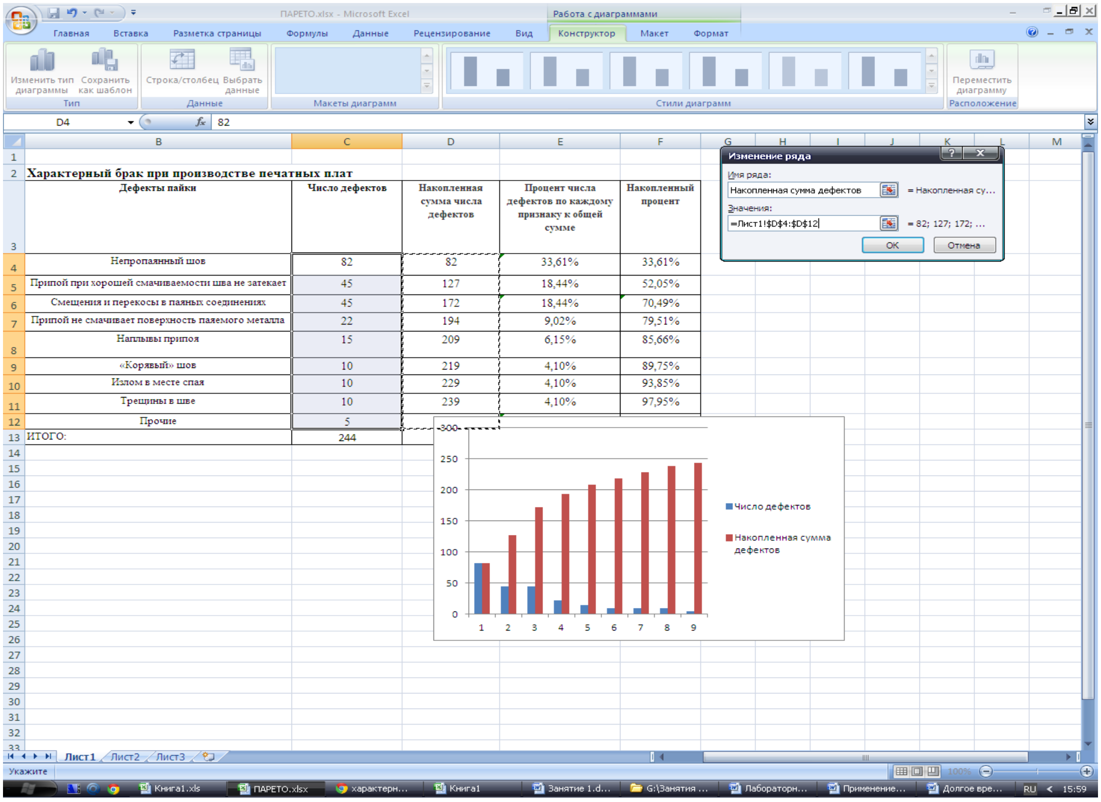


Рис. 81

Выделите ряд «Накопленная сумма дефектов» и в открывшемся меню выберите пункт «Изменить тип диаграммы для ряда» (рис. 82).

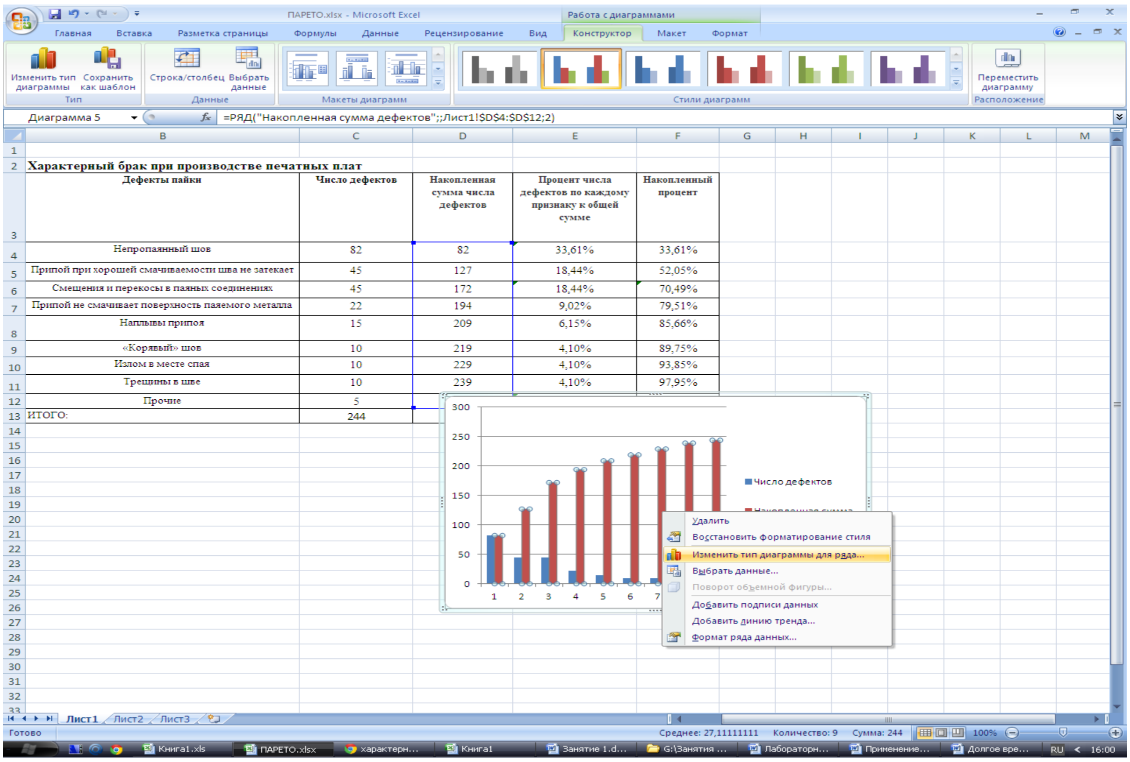


Рис. 82

В открывшемся диалоговом окне выберите тип графика. Нажмите ОК. Сделайте подписи легенды рядов. Добавьте подписи значений (рис. 83).

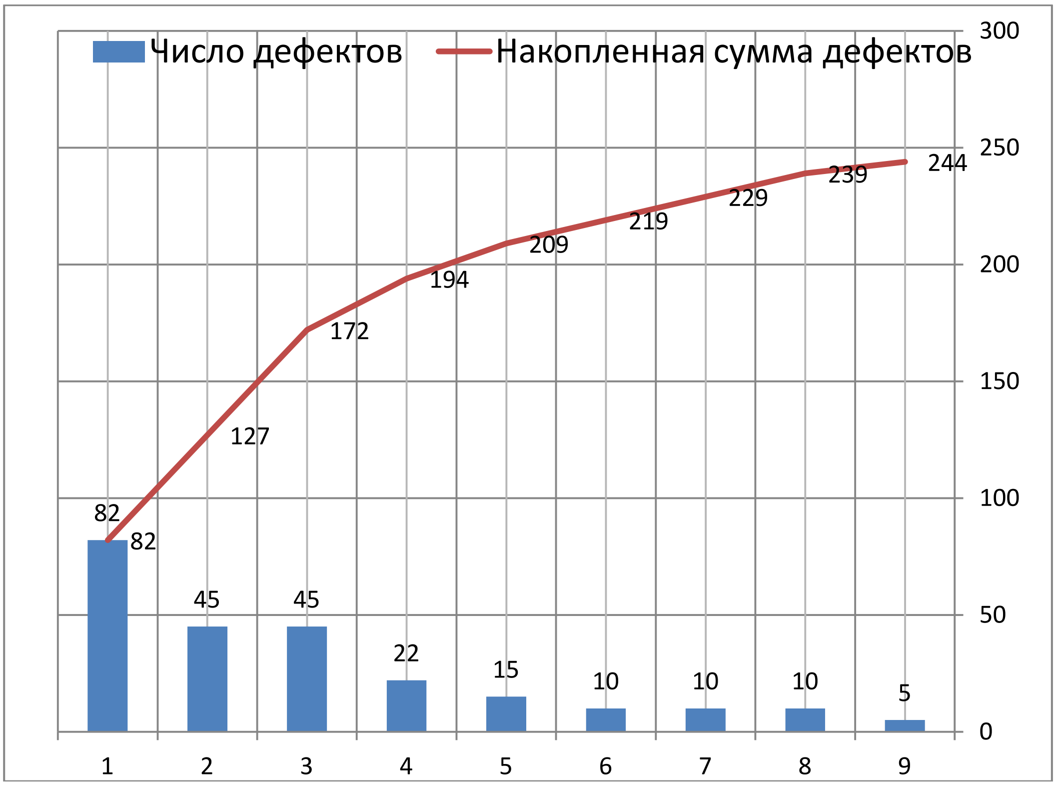


Рис. 83

Построение диаграммы Парето в STATISTICA 6.0

Для обработки данных выборки откройте файл MS Excel «Парето», сохраните его в формате Книга Excel 97-2003 и откройте сохраненный файл в программе Statistica 6.0. Для этого настройте схему импорта данных Листа 1 следующим образом (рис. 84).

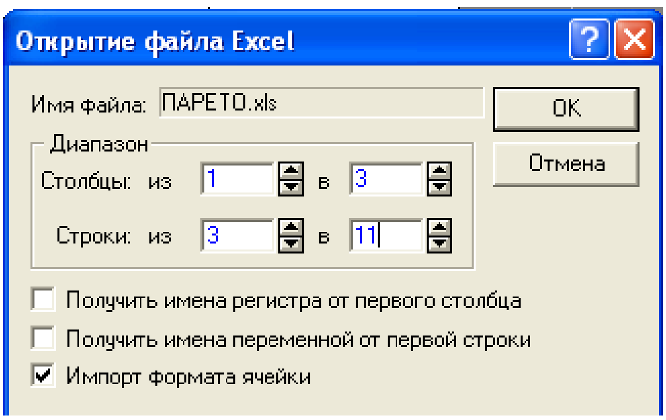


Рис. 84

Задайте заголовки переменных. Для этого щелкните по заголовку и измените название переменной (рис. 85).

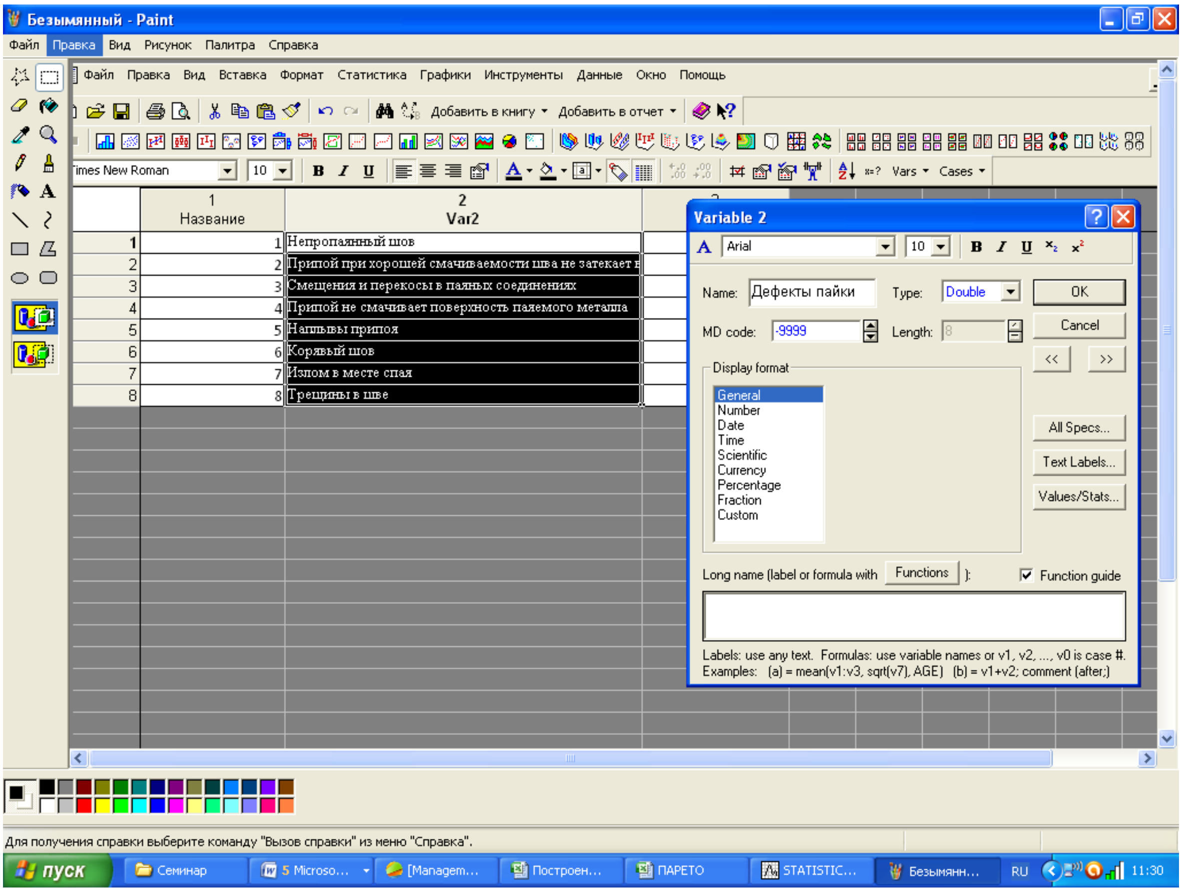


Рис. 85

Выберите в меню **Статистика-Индустриальная статистика&Сигма шесть-Качество диаграммы управления** (рис. 86).

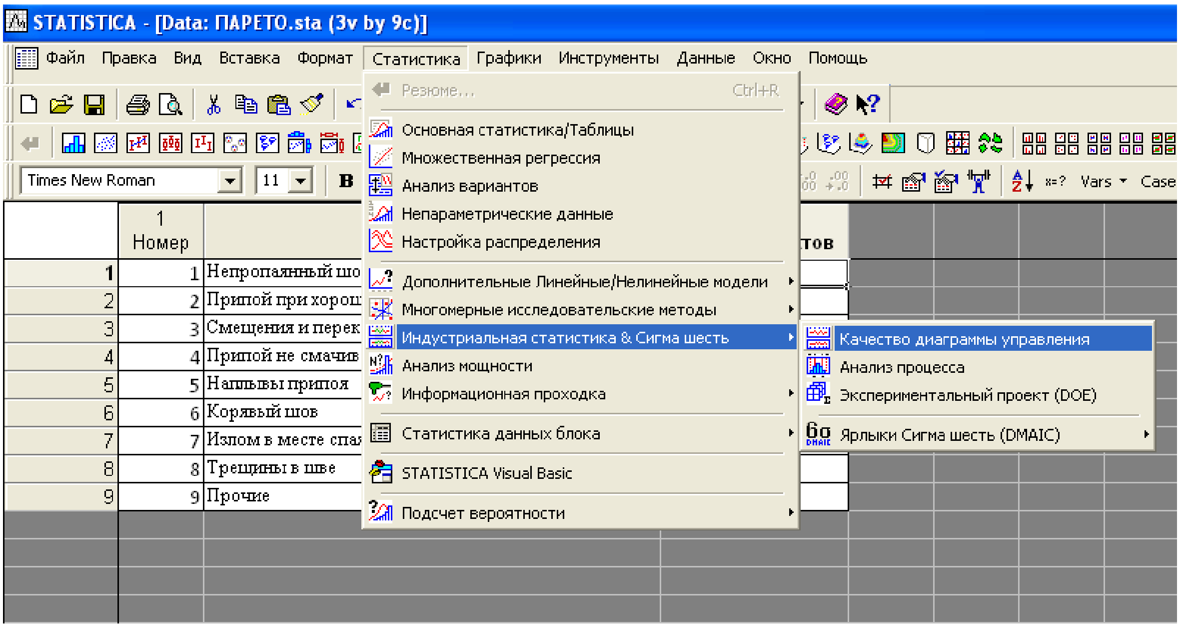


Рис. 86

В открывшемся списке выберите **Pareto chart analysis**. Нажмите **ОК** (рис. 87).

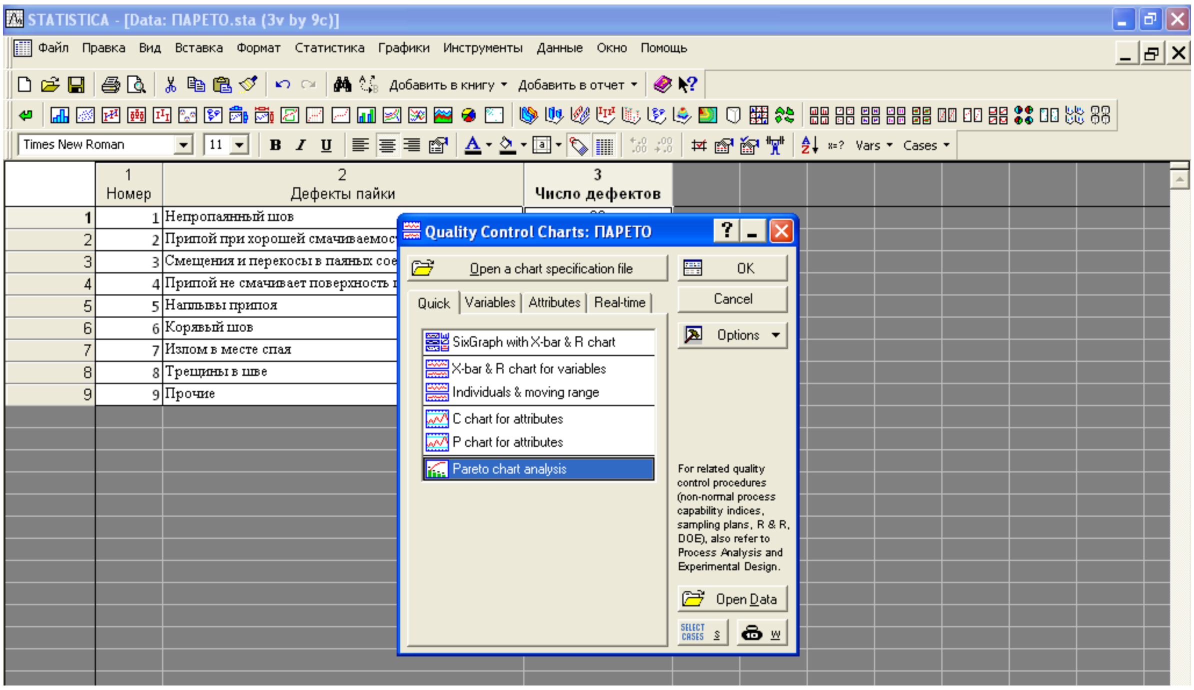


Рис. 87

Задайте переменную **Число дефектов**. Нажмите **ОК** (рис. 88).

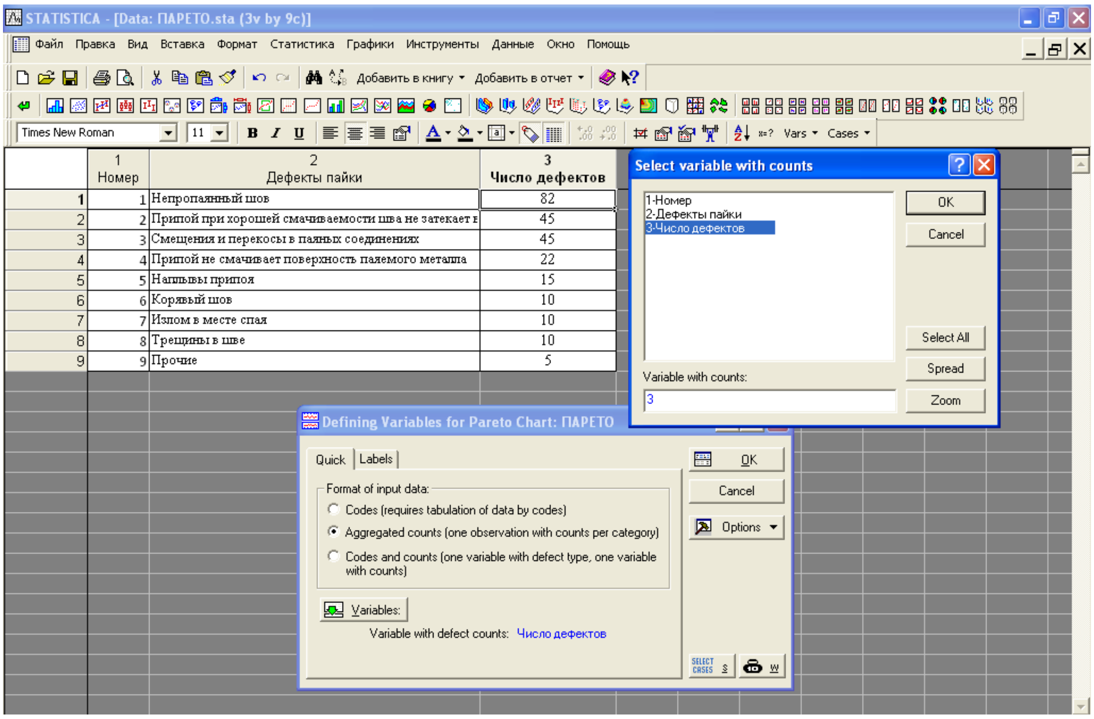


Рис. 88

На вкладке Labels выберите переменную **Номер**. Нажмите **ОК**. Для формирования диаграммы Парето нажмите **ОК** (рис. 89).

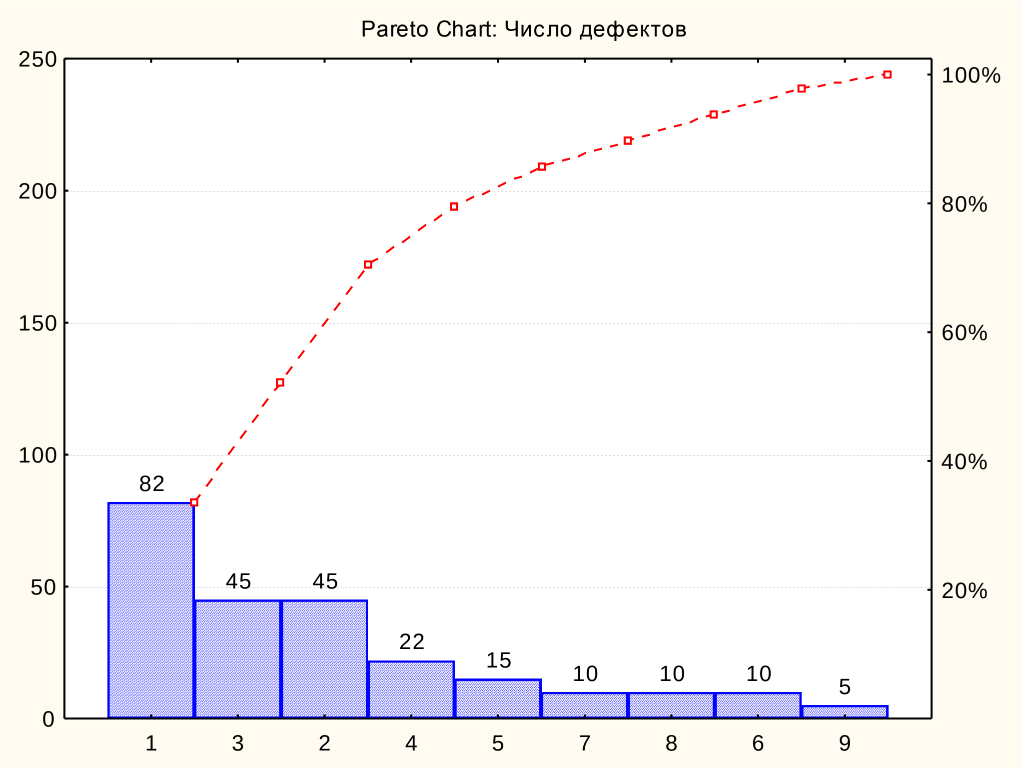


Рис. 89

ДИАГРАММА ИШИКАВЫ

Для выявления основных причин потери качества применятся достаточно простой, но эффективный инструменты аналитической визуализации – диаграмма Ишикавы (диаграмма причин и следствий, «рыбий скелет»).

Качественно составленная диаграмма Иш6икавы наглядно демонстрирует все «дыры» технологического процесса и способы их устранения. В виде больших костей отображаются причины, непосредственно влияющие на нарушение общего параметра (например, качества технологического процесса), а в виде малых костей – факторы, влияющие на причины. За счет управления факторами, можно влиять на причины и, в итоге, на конечный результат.

Существуют два вида диаграмм Ишикавы. Диаграмма типа 5М рассматривает такие компоненты качества, как "человек", "машина", "материал", "метод", "контроль", а в диаграмме типа 6М к ним добавляется компонент "среда". Применительно к решаемой задаче анализа для компоненты "человек" необходимо определить факторы, связанные с удобством и безопасностью выполнения операций; для компоненты "машина" — взаимоотношения элементов конструкции анализируемого изделия между собой, связанные с выполнением данной операции; для компоненты "метод" — факторы, связанные с производительностью и точностью выполняемой операции; для компоненты "материал" — факторы, связанные с отсутствием изменений свойств материалов изделия в процессе выполнения данной операции; для компоненты "контроль" — факторы, связанные с достоверным распознаванием ошибки процесса выполнения операции; для компоненты "среда" — факторы, связанные с воздействием среды на изделие и изделия на среду (рис. 90).

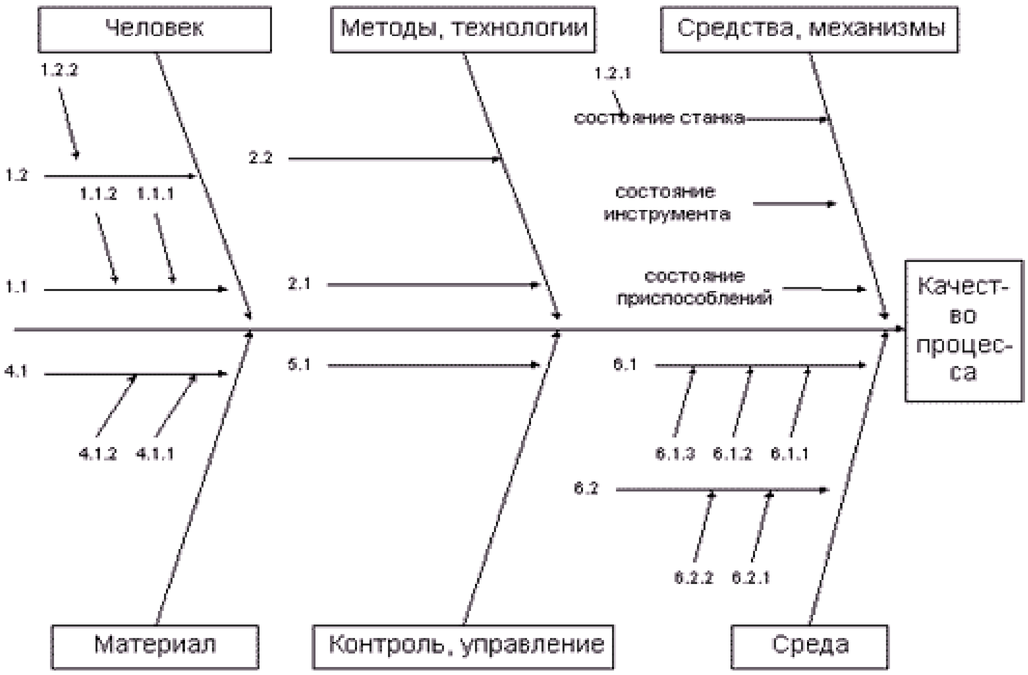


Рис. 90

Построение диаграммы Ишикавы в STATISTICA 6.0

Диаграмму Ишикавы легко построить при помощи инструментов пакета STATISTICA 6.0. Для этого достаточно правильно организовать хранение соответствующих факторов и причин в ячейках таблицы.

Рассмотрим причины выявленных дефектов пайки.

Непропаянный шов: Плохая зачистка места спая. Паяние производилось недостаточно нагретым паяльником.

Припой при хорошей смачиваемости шва не затекает в зазор: Мал зазор между спаиваемыми заготовками.

Смещения и перекосы в паяных соединениях: Некачественная фиксация взаимного положения заготовок перед пайкой.

Припой не смачивает поверхность паяемого металла: Недостаточная активность флюса. Наличие на поверхности оксидной пленки, жировых или других загрязнений.

Наплывы припоя: Использовано слишком обильное количество припоя.

Корявый шов: Паяние производилось недостаточно нагретым паяльником.

Излом в месте спая: Непропай шва.

Трещины в шве: Значительная разница в коэффициентах теплового расширения припоя и материала соединяемых частей.

Используемый метод контроля - технический осмотр изделия невооруженным глазом или с применением лупы в сочетании с измерениями позволяет проверить качество поверхности, заполнение зазоров припоем, наличие трещин и других наружных дефектов.

К вредным и опасным факторам в паяльном производстве относятся: ультрафиолетовое видимое и инфракрасное излучение источников нагрева и нагретых деталей; запыленность и загазованность воздуха. При пайке, напылении, выплавке припоев и флюсов в окружающую среду поступают аэрозоли, содержащие в составе твердой фазы окислы металлов (марганца, хрома, никеля, железа, меди, титана, алюминия), а также токсичные газы (окись углерода, фтористые, хлористые, бромистые соединения, окислы азота). В составе аэрозолей могут быть составляющие флюсов и припоев, содержащих свинец, кадмий, цинк, олово, углеводороды. Количество аэрозолей, их токсичность зависят от состава припоев, флюсов, технологии и степени механизации производства.

Рассмотренные данные сведем в таблицу 21 и внесем в качестве переменных в программу STATISTICA 6.0 (рис. 91).

Таблица 21

"человек"	**"машина"**	**"материал"**	**"метод"**	**"контроль"**	**"среда"**
Плохая зачистка	Температура паяльника	Активность флюса	Зазор для спаивания	Технический осмотр изделия	ультрафиолетовое излучение
Излишек припоя		Коэффициенты теплового расширения			инфракрасное излучение
Непропай шва					окислы металлов
Наличие загрязнений					токсичные газы
Некачественная фиксация					запыленность воздуха
					загазованность воздуха

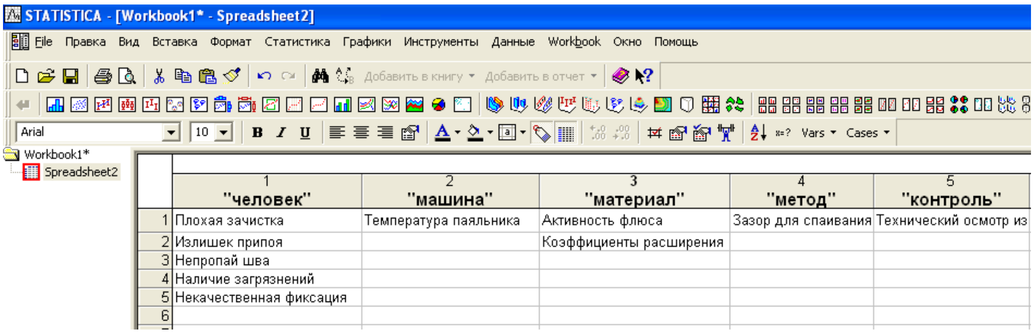


Рис. 91

Выберите в меню Статистика-Индустриальная статистика&Сигма шесть-Анализ процесса(рис. 92).

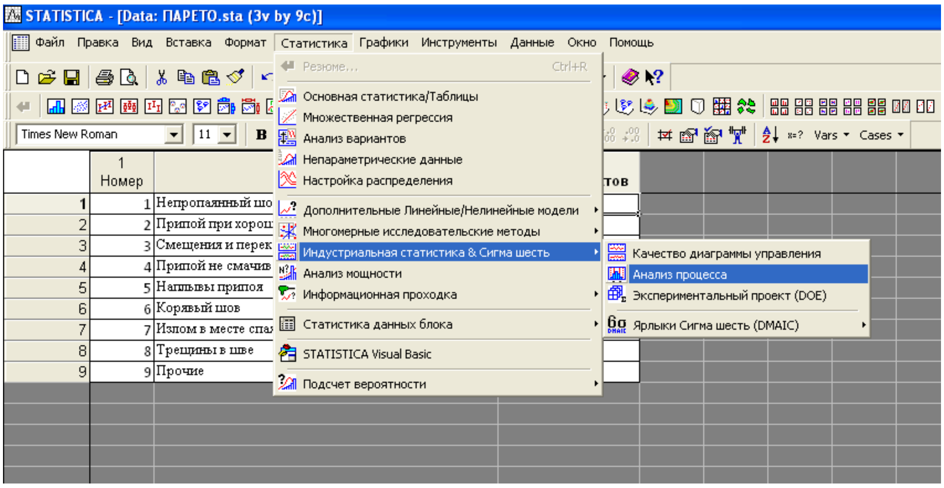


Рис. 92

В открывшемся списке выберите **Cause-effect (Ishikawa, Fishbone) diagrams**. Нажмите **ОК** (рис. 93).

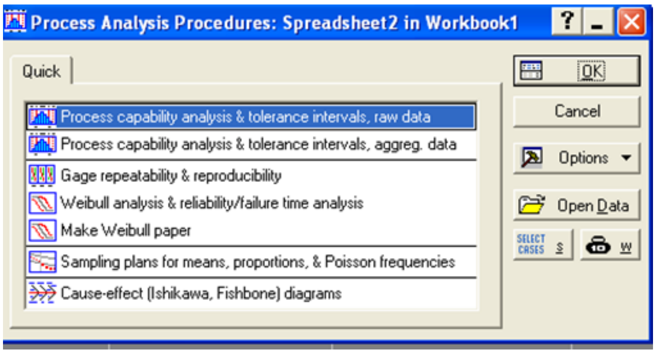


Рис. 93

Нажмите на кнопку задания переменных, после чего в открывшемся окне выберите данные.В открывшемся окне выберите расположение данных по отношению к центральной линии. Нажмите **ОК** (рис. 94).

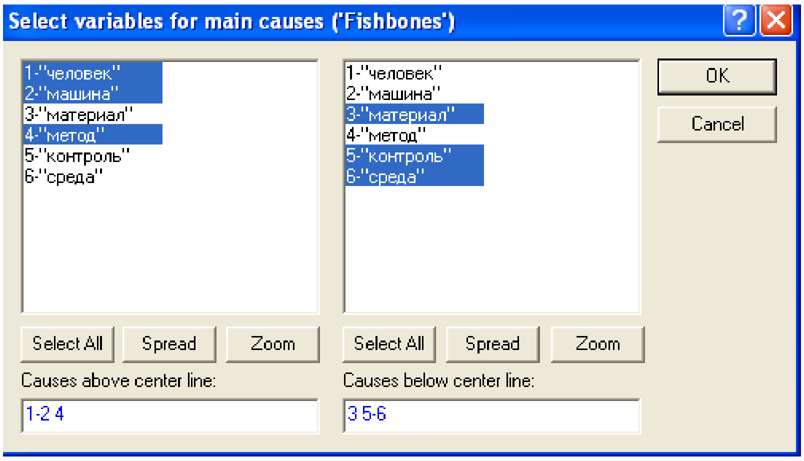


Рис. 94

Нажмите на кнопку построения диаграммы Ишикавы. Для получения желаемого вида диаграммы используйте возможности настройки двумерных графиков программы, например, выделите заголовок, введите новый текст. Если на диаграмме потребуется изображение «малых костей», их нужно нарисовать и сделать к ним надписи (рис. 95).

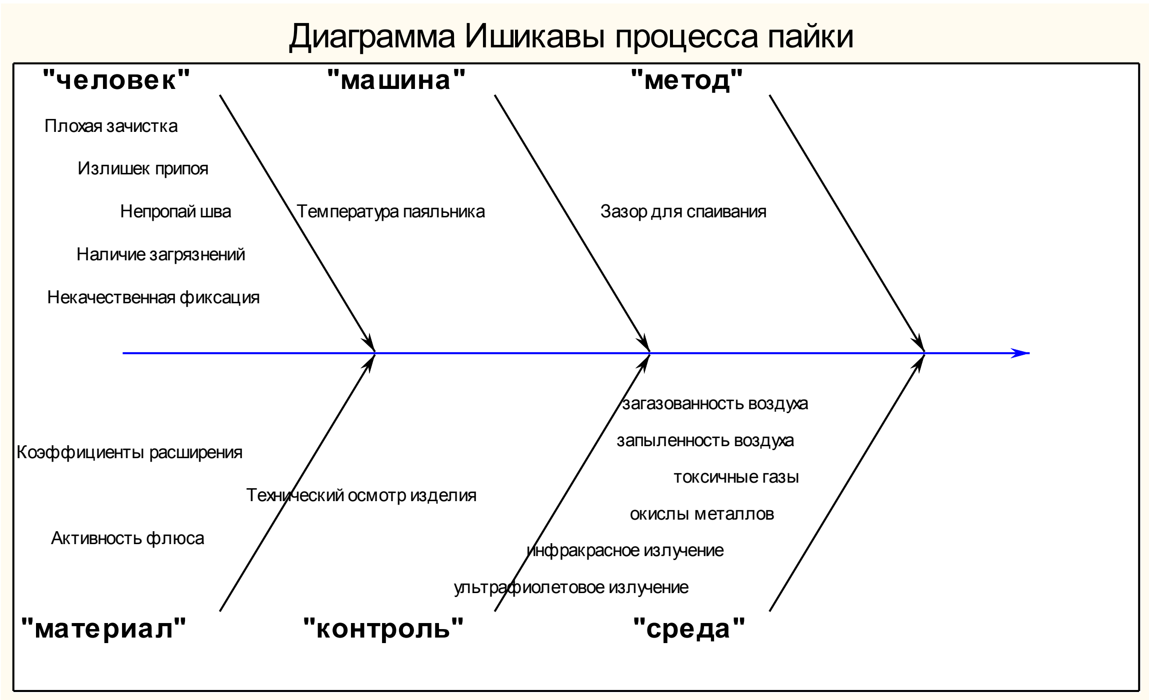


Рис. 95

После построения диаграммы Ишикавы можно построить диаграмму Парето по причинам.

ДИАГРАММА РАЗБРОСА

Метод "Диаграмма разброса" - один из инструментов статистического контроля качества применяется в производстве и на различных стадиях жизненного цикла продукции для выяснения зависимости между показателями качества и основными факторами производства. Достоинства метода - наглядность и простота оценки связей между двумя переменными.

Японский союз ученых и инженеров в 1979 г. включил диаграмму разброса в состав семи методов контроля качества.

Цель метода - выяснение существования зависимости и выявление характера связи между двумя различными параметрами процесса.

Диаграмма разброса - инструмент, позволяющий определить вид и тесноту связи между парами соответствующих переменных. Эти две переменные могут относиться:

- к характеристике качества и влияющему на нее фактору;
- двум различным характеристикам качества;
- двум факторам, влияющим на одну характеристику качества.

При наличии корреляционной зависимости между двумя факторами значительно облегчается контроль процесса.

Диаграмма разброса в процессе контроля качества используется также для выявления причинно-следственных связей показателей качества и влияющих факторов.

Диаграмма разброса - это точечная диаграмма в виде графика, получаемого путем нанесения в определенном масштабе экспериментальных, полученных в результате наблюдений точек. Координаты точек на графике соответствуют значениям рассматриваемой величины и влияющего на него фактора. Расположение точек показывает наличие и характер связи между двумя переменными (например, скорость и расход бензина, или выработанные часы и выход продукции).

По полученным экспериментальным точкам могут быть определены и числовые характеристики связи между рассматриваемыми случайными величинами: коэффициент корреляции и коэффициенты регрессии. На рисунке96 представлены виды диаграммы разброса.

Рис. 96

Правила построения диаграммы разброса

1) Определить, между какими парами данных необходимо установить наличие и характер связи. Желательно не менее 25-30 пар данных.

2) Для сбора данных подготовить бланк таблицы (листок регистрации), предусмотрев в нем графы для порядкового номер наблюдения i; независимой переменной характеристики, называемой аргументом x; зависимой переменной, называемой функцией (откликом) y.

3) По результатам наблюдения заполнить листок регистрации данных.

4) По полученным данным построить график в координатах x-y и нанести на него данные. Длина осей, равная разности между максимальными и минимальными значениями для x и y, по вертикали и по горизонтали должна быть примерно одинаковой, тогда диаграмму будет легче читать.

5) Нанести на диаграмму все необходимые обозначения. Данные, отраженные на диаграмме, должны быть понятны любому человеку, а не только тому, кто делал диаграмму.

Для выяснения влияния одной переменной на другую следует собрать необходимые данные и внести их в листок регистрации.

По полученным данным построить диаграмму разброса и провести анализ диаграммы. Иногда желательно получить количественную оценку тесноты или силы связи между случайными величинами – коэффициент корреляции.

Что же такое коэффициент корреляции?

Предположим, что мы измеряем две величины x и y, характеризующие два различных экспериментальных образца под номерами 1 и 2. Обозначим результаты измерений на первом образце (x_1, y_1), а на втором - (x_2, y_2).

Можно проделать эти операции для любого числа образцов, например n и получить набор n парных наблюдений (x_1, y_1), (x_2, y_2),..., (x_n, y_n).

Обозначим среднее значение величин x- $\bar{x}$, а y-$\bar{y}$ и рассчитаем отклонения каждой пары наблюдений от их средних значений. Так для i -го образца они равны:

$$(x_i - \bar{x}, y_i - \bar{y}).$$

Эти отклонения можно представить n точками на диаграмме: пересечению осей соответствует точка $(\bar{x}, \bar{y})$- нуль отклонений. Проделав это для конкретного множества данных, получим картину рассеяния данных (рис. 97).

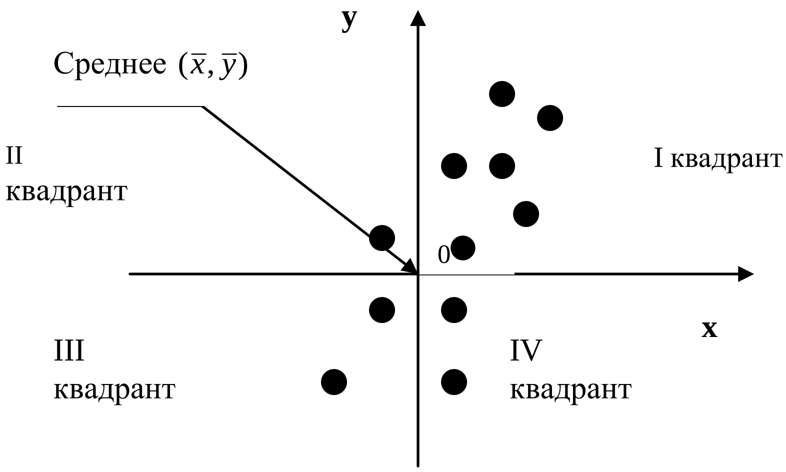


Рис. 97

Оси делят плоскость на четыре квадранта, занумерованных, как показано на рисунке. Большинство наблюдений лежит в I и III квадрантах, и лишь немногие — во II и IV. Наблюдения в I и III квадрантах соответствуют значениям *x* выше среднего, связанным со значениями *y* выше среднего, и значениям *x* ниже среднего, связанным со значениями *y* ниже среднего.

Удобно ввести сокращенные обозначения для отклонений величин от средних значений. Отклонения для i -го образца запишутся в виде X_i и Y_i, т.е.

$$X_i = x_i - \bar{x} \tag{61}$$

$$Y_i = y_i - \bar{y} \tag{62}$$

Для всех точек из I квадранта X_i и Y_i- положительны, и потому их произведение - также положительно.

В III квадранте X_i и Y_i- отрицательны, так что их произведение снова положительно.

Для тех немногих точек, которые лежат во II и IV квадранте, одно из отклонений положительно, а другое отрицательно, и потому их произведение отрицательно.

Если сложить эти произведения для всех n и обозначить результат через S_{xy} получим:

$$S_{xy} = X_1Y_1 + X_2Y_2 + ... + X_nY_n\text{, или } S_{xy} = \sum_{i=1}^{n} X_iY_i \qquad (63)$$

Величина S_{xy} называется *суммой произведений отклонений от среднего.* Если точки попадают преимущественно в I и III квадранты, сумма велика и положительна, так как положительно большинство слагаемых. Если большинство точек лежит во II и IV квадрантах, сумма велика и отрицательна. Если же точки рассеяны равномерно по всем квадрантам, то слагаемые стремятся взаимно сократить друг друга, и сумма оказывается близкой к нулю. Ясно, что значение S_{xy} зависит как от значения n, так и от «разброса» наблюдаемых значении *x* и *y.*

Чтобы получить меру этой связи, независимую от разброса и числа наблюдении, разделим $S_{xy} = X_1Y_1 + X_2Y_2 + ... + X_nY_n$ на меру разброса, которая сама растет с числом наблюдений. Обычно, в качестве меры разброса, берут меру, построенную на сумме квадратов отклонений значений *x* и *y* от их средних, т. е. $\sum_{i=1}^{n} X_i^2$ и $\sum_{i=1}^{n} Y_i^2$.

Мы же фактически будем делить на корень квадратный из произведения этих двух мер. Возникающая в результате величина, которую обычно обозначают буквой r, называется *коэффициентом корреляции.*

Он рассчитывается следующим образом:

$$r = \frac{\sum_{i=1}^{n} X_iY_i}{\sqrt{\sum_{i=1}^{n} X_i^2 \sum_{i=1}^{n} Y_i^2}} \qquad (64)$$

Можно показать теоретически, что если между величинами x и y существует точно линейная зависимость, r принимает значение +1 или –1:

- +1, если линия идет вверх слева направо;
- - 1, если линия идет вниз слева направо.

Если имеет место полное рассеяние, т.е. если связи между величинами x и y нет, r принимает значение близкое к нулю.

Построение диаграммы разброса в MS Office Excel

Построение и анализ диаграммы рассеяния произведем на данных микробиологического лабораторного контроля показателя общей бактериальной обсемененности (КМАФАнМ) сливок после операции пастеризации в процессе получения сливочного масла на молочном заводе.

Общая бактериальная обсемененность (КМАФАнМ) - количество мезофильных аэробных и факультативно-анаэробных микроорганизмов в 1 г или 1 см^3 продукта. Высокая бактериальная обсемененность пищевых продуктов свидетельствует о недостаточной термической обработке сырья в процессе пастеризации. Нормы КМАФАнМопределены в СанПиН 2.3.2.560-96, выписка из которого для выпускаемых видов масла сливочного приведена в таблице 22.

Таблица 22

Вид масла	КМАФАнМ,КОЕ/г (см^3),не более
Масло вологодское	$1x10^4$
Масло сладко-сливочное и соленое любительское и крестьянское	$1x10^5$
Масло кисло-сливочное любительское и крестьянское	-
Масло шоколадное	$1x10^5$
Масло сливочное бутербродное	$5x10^5$

В таблице 23 представлены результаты контроля показателей процесса производства сливочного масла – температура процесса пастеризации и значение показателя (КМАФАнМ) после пастеризации.

Таблица 23

Номер	Температура пастеризации ^{0}C X_i	КМАФАнМ коЕ/см^3 Y_i
1	99,5	100
2	103,0	300
3	96,4	870
4	99,0	500
5	99,0	500
6	97,4	1700
7	96,0	2400
8	100,4	440
9	102,0	1000
10	97,0	6200
11	102,0	280
12	100,5	460
13	96,2	2600
14	99,8	260
15	99,0	390
16	97,2	1800
17	96,0	930
18	100,1	580
19	102,3	890
20	97,1	5900
21	99,2	102
22	102,1	940
23	96,1	930
24	99,6	380
25	99,1	480
26	97,4	1600
27	96,0	3100
28	100,5	450
29	99,5	490
30	97,0	530

По данным таблицы построим диаграмму рассеяния – точечный график (рис. 98).

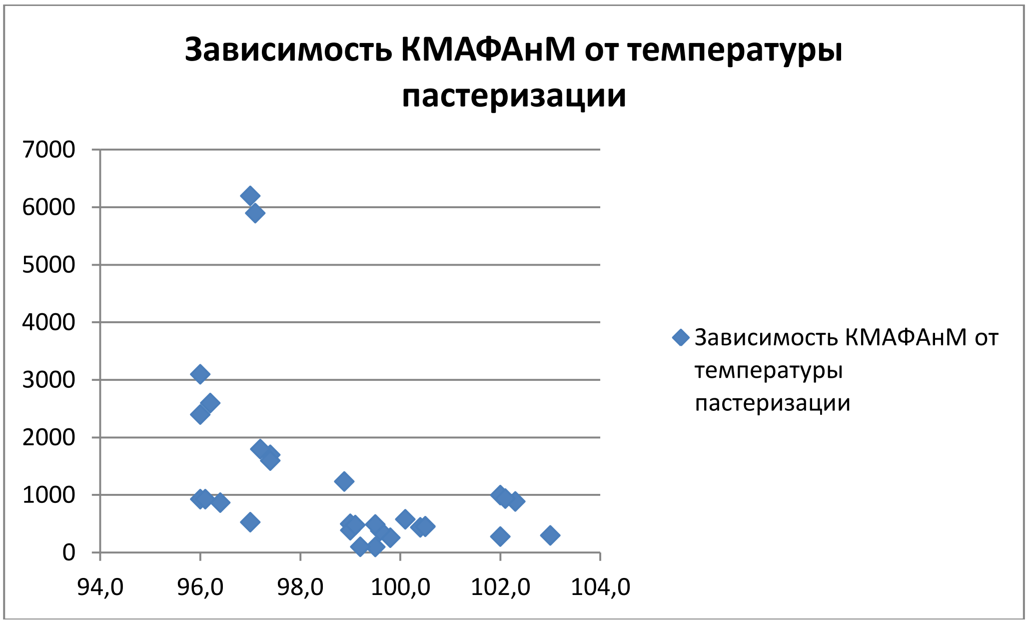


Рис. 98

По полученным данным рассчитаем среднее значение температуры пастеризации -98,88^{0}С и среднее значение КМАФАнМ 1236,73 коЕ/см^3

Расчет коэффициента корреляции произведите в таблице 24, или с помощью статистической функции путем ввода в свободную ячейку формулы:

=КОРРЕЛ(D2:D31;E2:E31).

Таблица 24

$X_i = x_i - \bar{x}$	$Y_i = y_i - \bar{y}$	$S_{xy} = \sum_{i=1}^{n} X_i Y_i$	$\sum_{i=1}^{n} X_i^2$	$\sum_{i=1}^{n} Y_i^2$
0,62	-1136,73	-704,77	0,38	1292162,67
4,12	-936,73	-3859,34	16,97	877469,34
-2,48	-366,73	909,50	6,15	134493,34
0,12	-736,73	-88,41	0,01	542776,00
0,12	-736,73	-88,41	0,01	542776,00
-1,48	463,27	-685,63	2,19	214616,00
-2,88	1163,27	-3350,21	8,29	1353189,34
1,52	-796,73	-1211,03	2,31	634784,00
3,12	-236,73	-738,61	9,73	56042,67
-1,88	4963,27	-9330,94	3,53	24634016,00
3,12	-956,73	-2985,01	9,73	915338,67
1,62	-776,73	-1258,31	2,62	603314,67
-2,68	1363,27	-3653,55	7,18	1858496,00
0,92	-976,73	-898,59	0,85	954008,00
0,12	-846,73	-101,61	0,01	716957,34
-1,68	563,27	-946,29	2,82	317269,34

-2,88	-306,73	883,39	8,29	94085,34
1,22	-656,73	-801,21	1,49	431298,67
3,42	-346,73	-1185,83	11,70	120224,00
-1,78	4663,27	-8300,61	3,17	21746056,00
0,32	-1134,73	-363,11	0,10	1287619,74
3,22	-296,73	-955,48	10,37	88050,67
-2,78	-306,73	852,72	7,73	94085,34
0,72	-856,73	-616,85	0,52	733992,00
0,22	-756,73	-166,48	0,05	572645,34
-1,48	363,27	-537,63	2,19	131962,67
-2,88	1863,27	-5366,21	8,29	3471762,67
1,62	-786,73	-1274,51	2,62	618949,34
0,62	-746,73	-462,97	0,38	557610,67
-1,88	-706,73	1328,66	3,53	499472,00
ИТОГО:		-45957,36	133,27	66095523,87

Выберите свободную ячейку и введите в нее **формулу:** =H32/КОРЕНЬ(I32*J32). В результате вычислений получено значение r=-0,49.

Понятие об эмпирических формулах. Метод наименьших квадратов

На практике мы часто сталкиваемся с задачей о сглаживании экспериментальных зависимостей.

Пусть зависимость между двумя переменными x и y выражается в виде таблицы, полученной опытным путем. Это могут быть результаты опыта или наблюдений, статистической обработки материала и т.п.

X	X_1	X_2	…	X_i	…	X_n
Y	Y_1	Y_2	…	Y_i	…	n

Требуется наилучшим образом сгладить экспериментальную зависимость между переменными **X** и **Y** , т.е. по возможности точно отразить общую тенденцию зависимости **Y** от **X**, исключив при этом случайные отклонения, связанные с неизбежными погрешностями измерений или статистических наблюдений. Такую сглаженную зависимость стремятся представить в виде формулы Y=F(X).
Формулы, служащие для аналитического представления опытных данных, получили название ***эмпирических формул***.

Задача нахождения эмпирических формул разбивается на два этапа. На первом этапе нужно установить **вид зависимости** Y=F(X), т.е. решить, является ли она линейной, квадратичной, логарифмической или какой-либо другой.
На первом этапе строят диаграмму рассеяния - на плоскость наносят точки, являющиеся результатами экспериментальных исследований и выдвигают гипотезу о виде зависимости.

Предположим, первый этап завершен — вид функции Y=F(X) установлен. Тогда переходят ко второму этапу — **определению неизвестных параметров этой функции**.

Согласно наиболее распространенному и теоретически обоснованному методу наименьших квадратов в качестве неизвестных параметров функции F(X) выбирают такие значения, чтобы сумма квадратов невязок δ_i , или отклонений *"теоретических" значении* F(X_i), найденных по эмпирической формуле Y=F(X) от соответствующих *опытных значений* Y_i, была минимальной.

Уравнение линейной регрессии

В 1886 г. генетик Фрэнсис Гальтон заметил, что родители большого роста имеют преимущественно и детей большого роста, а родители небольшого роста — детей небольшого роста. Когда Гальтон нанес на график средний рост старших сыновей для различных значений среднего роста родителей, он получил почти прямую линию, проходящую через нанесенные точки.

Поскольку рост потомства стремится двигаться к среднему, Гальтон назвал это явление *регрессией к* среднему, а линию, проходящую через точки на графике,— *линией регрессии.* Термин«линия регрессии» употребляется теперь для *линий наилучшей подгонки* под экспериментальные точки вне зависимости от того, имеется ли регрессия к среднему в смысле Гальтона.

Пусть в качестве эмпирической формулы функции взята ***линейная функция*** $y = ax + b$

Задача сводится к отысканию таких значений параметров *a* и *b*, при которых функция

$$S = \sum_{i=1}^{n} (ax_i + b - y_i)^2 \qquad (65)$$

принимает наименьшее значение. Функция $S = S(a;b)$ есть функция двух переменных a и b до тех пор, пока не найдены их «наилучшие» (в смысле метода наименьших квадратов) значения. Следовательно, для нахождения прямой, наилучшим образом согласованной с опытными данными достаточно решить систему, приравняв нулю частные производные подбираемой линейной функции:

$$\begin{cases} S_a' = 0 \\ S_b' = 0 \end{cases}, \quad \text{или} \quad \begin{cases} \sum_{i=1}^{n} 2(ax_i + b - y_i)x_i = 0 \\ \sum_{i=1}^{n} 2(ax_i + b - y_i) = 0 \end{cases}$$

После алгебраических преобразований эта система принимает вид:

$$\begin{cases} (\sum_{i=1}^{n} x_i^2) \times a + (\sum_{i=1}^{n} x_i) \times b = \sum_{i=1}^{n} x_i y_i \\ (\sum_{i=1}^{n} x_i) \times a + n \times b = \sum_{i=1}^{n} y_i \end{cases}$$

Полученная система нормальных уравнений имеет единственное решение (a,b), соответствующее минимуму функции $S = \sum_{i=1}^{n} (a \times x_i + b - y_i)^2$.

Параметры a, b линейной модели в MS Excel можно определить при помощи функций НАКЛОН (SLOPE) и ОТРЕЗОК (INTERCEPT).

Функция НАКЛОН (SLOPE) определяет коэффициент наклона линейного тренда, а функция ОТРЕЗОК (INTERCEPT) – точку пересечения линии линейного тренда с осью ординат.

Синтаксис:

НАКЛОН (изв_знач_у; изв_знач_х),

ОТРЕЗОК (изв_знач_у; изв_знач_х),

где изв_знач_у – массив известных значений зависимой наблюдаемой величины;

изв_знач_х – массив известных значений независимой наблюдаемой величины. Если изв_знач_х опущены, то предполагается, что это массив {1,2,3;…} такого же размера, как изв_знач_у.

Постройте уравнение линейной регрессии для результатов контроля показателей процесса производства сливочного масла – температура процесса пастеризации и значение показателя (КМАФАнМ) после пастеризации. Для этого рассчитайте параметры уравнения линейной регрессии $y = ax + b$

Выберите свободную ячейку и введите в нее формулу расчета a: НАКЛОН НАКЛОН(D2:D32;C2:C32).

В другую свободную ячейку введите в нее формулу расчета b: ОТРЕЗОК(D2:D32;C2:C32).

В результате расчета получены следующие значения a= - 344, b = 35270, по которым построен график (рис. 99).

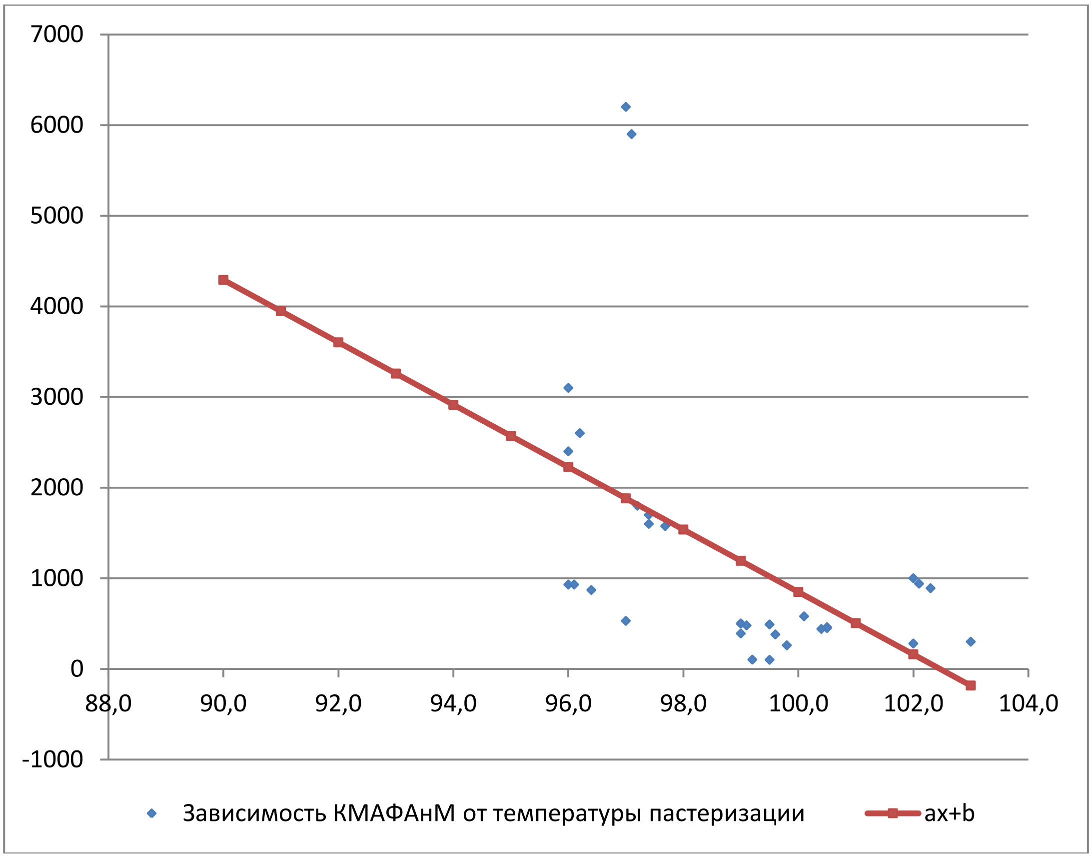


Рис. 99

Построение диаграммы разброса в STATISTICA 6.0

Для построения диаграммы разброса выберите в меню **Графики – Графики рассеяния** (рис. 100).

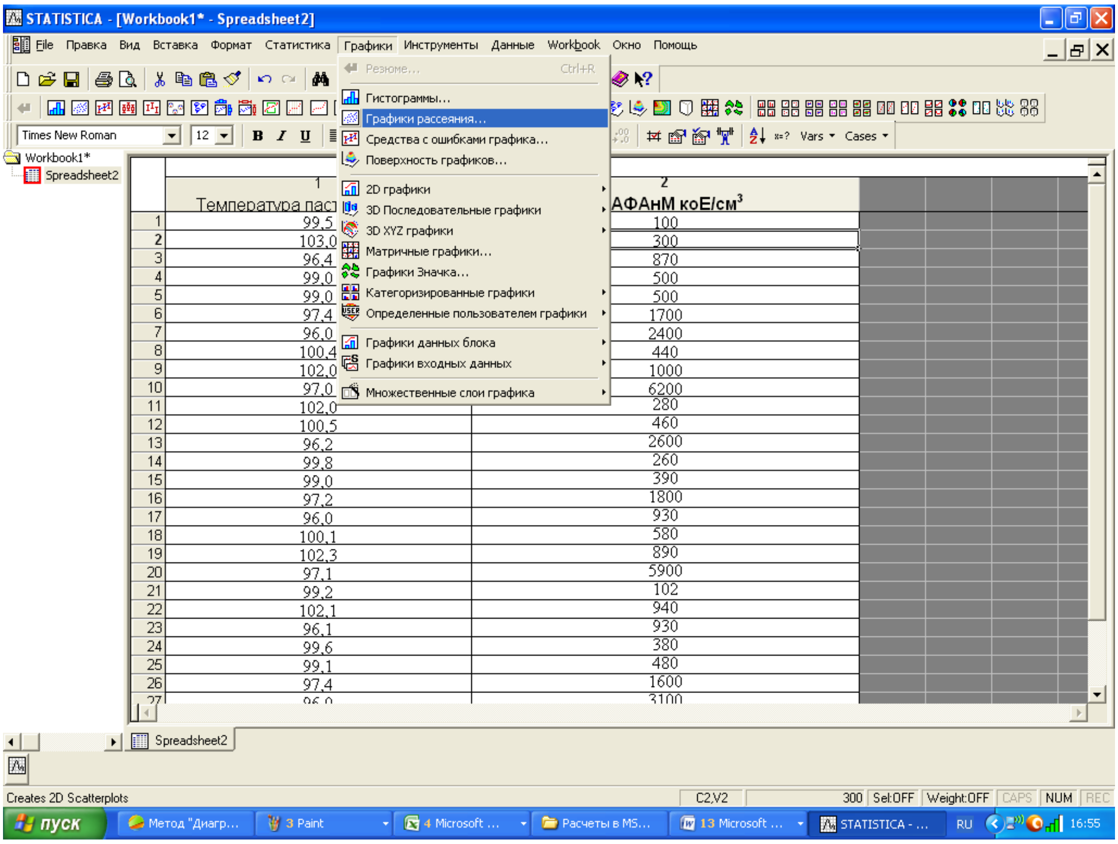


Рис. 100

Задайте переменные X_i и Y_i. Нажмите **ОК** (рис. 101).

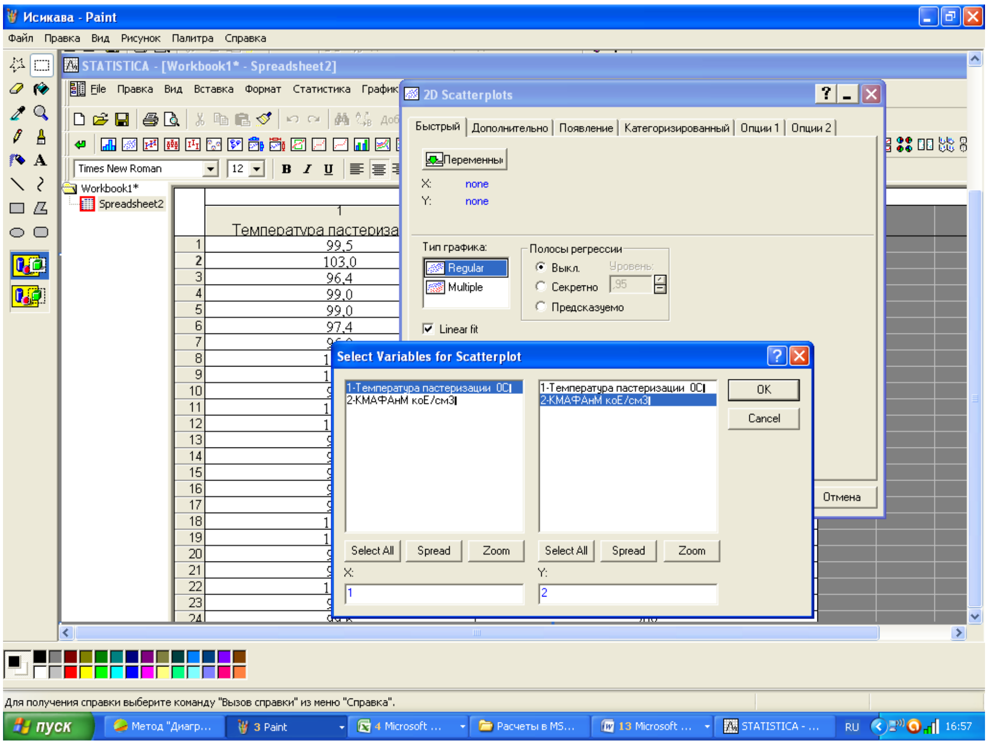


Рис. 101

Нажмите **ОК** (рис. 102).

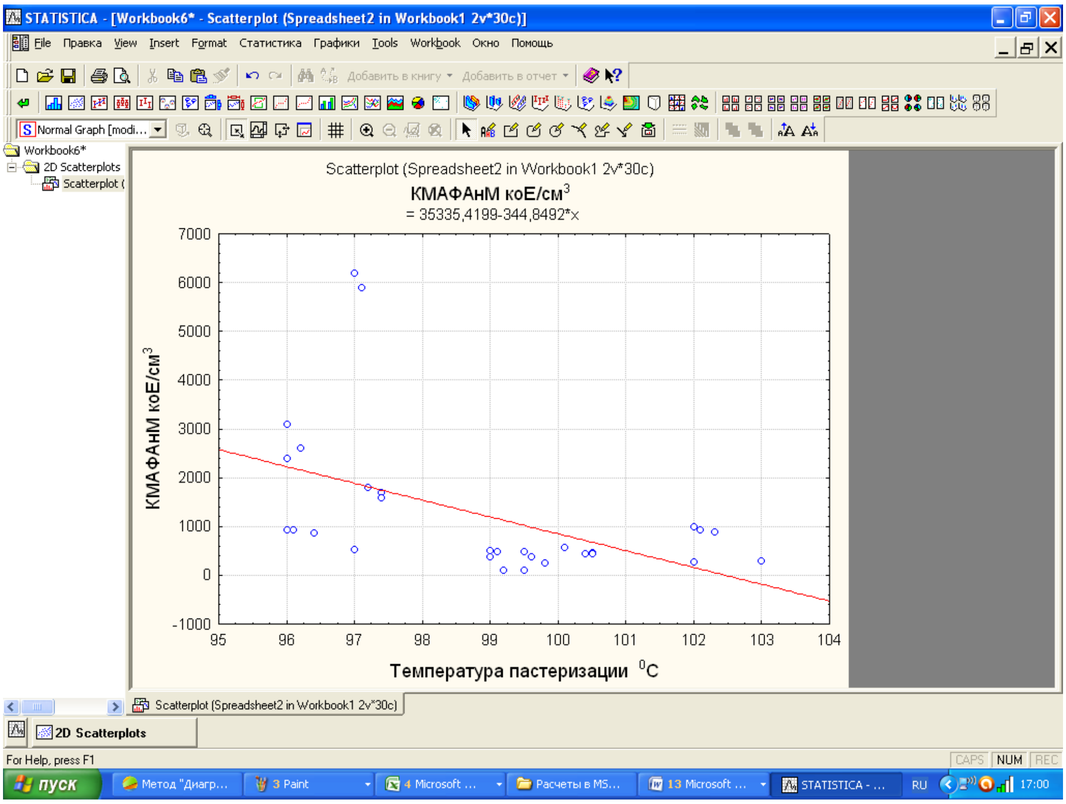


Рис. 102

Для проведения корреляционного анализа выберите в меню **Статистика – Основная статистика/Таблицы- Correlation matrices**(рис. 103).

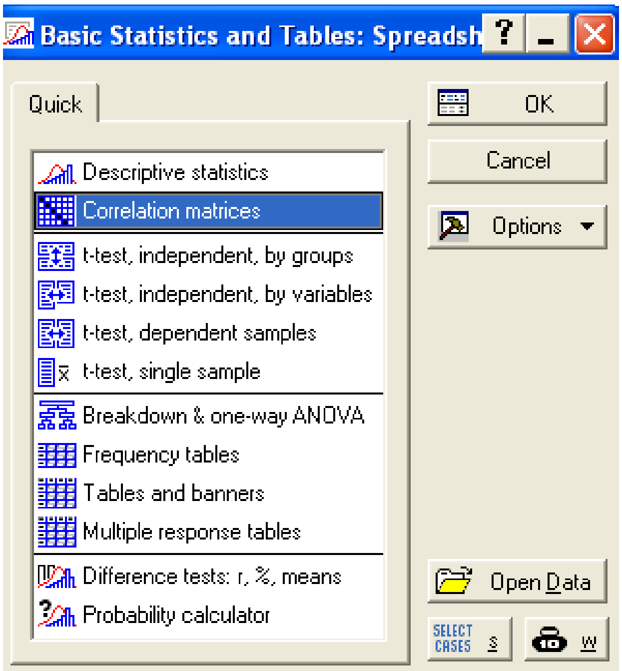


Рис. 103

Для выбора переменных нажмите на кнопку **Two lists (rect. Matrix),** задайте переменные и нажмите **ОК** (рис. 104).

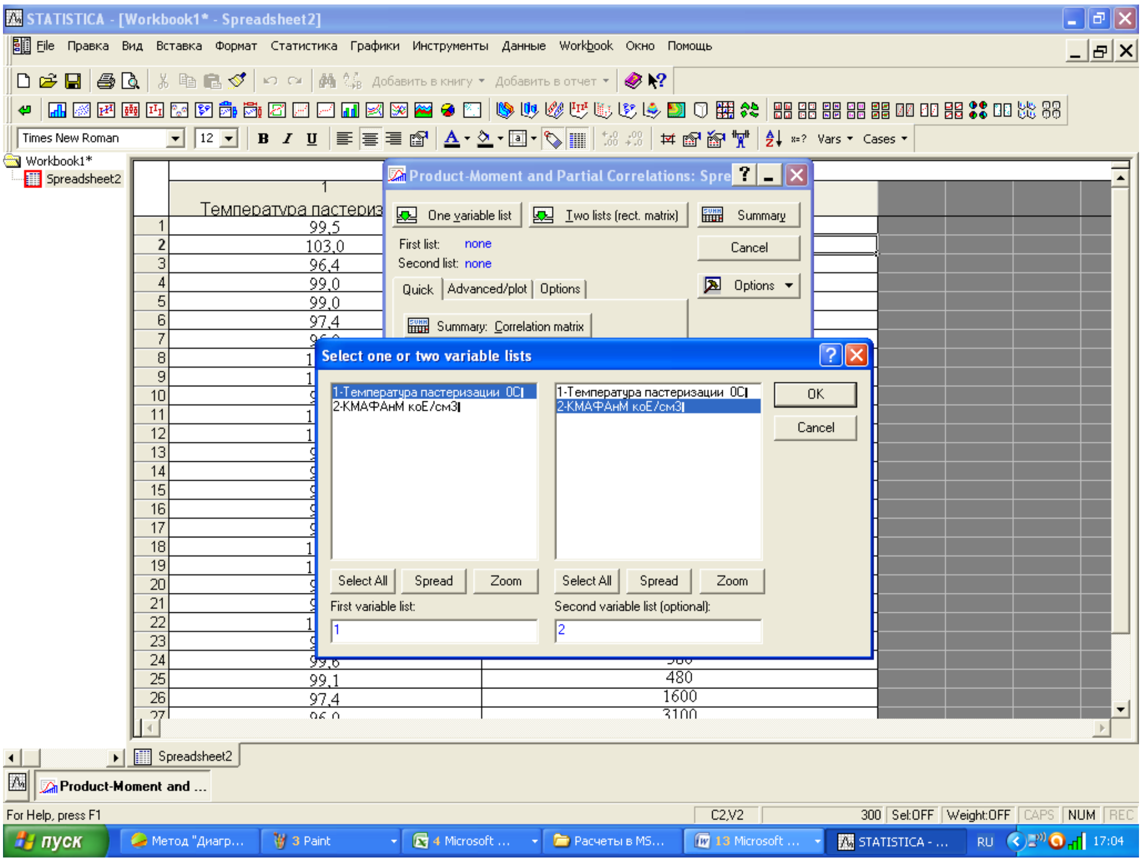


Рис. 104

После нажатия на соответствующие кнопки будет произведен расчет коэффициента корреляции и построена диаграмма, на которой можно увидеть гистограммы распределения переменных и диаграмму рассеяния (рис. 105).

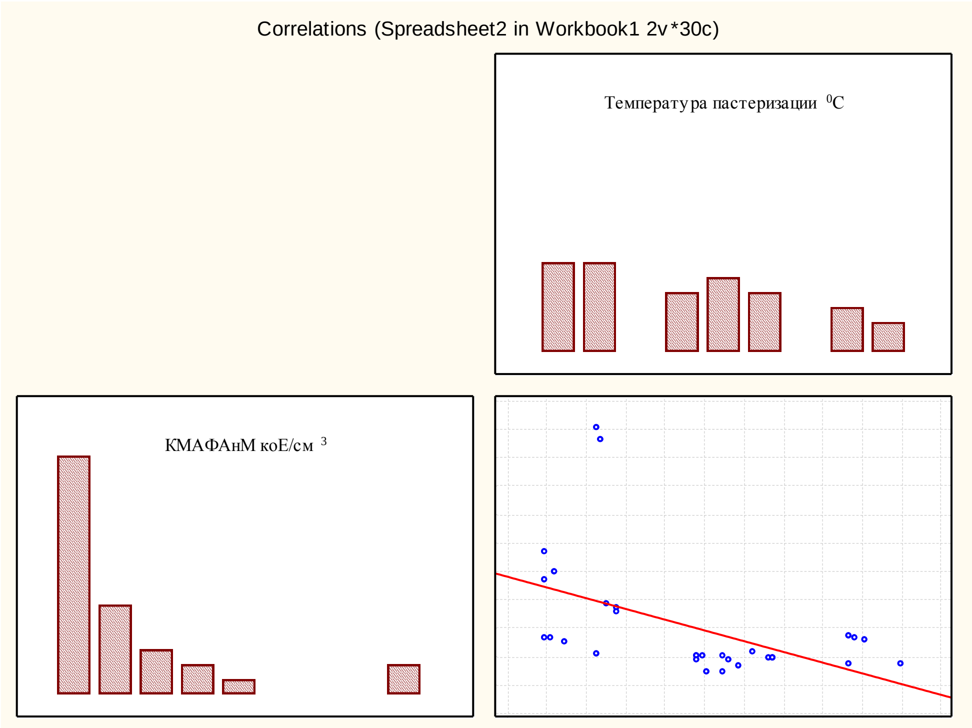


Рис. 105

СТРАТИФИКАЦИЯ

В основном, стратификация — процесс сортировки данных согласно некоторым критериям или переменным, результаты которого часто показываются в виде диаграмм и графиков. Можно классифицировать массив данных в различные группы (или категории) с общими характеристиками, называемыми переменной стратификации. Важно установить, которые переменные будут использоваться для сортировки. Стратификация — основа для других инструментов, таких как анализ Парето или диаграммы рассеивания. Такое сочетание инструментов делает их более мощными.

Стратификация данных выполняется следующим образом:

– Определяются факторы, по которым будет проводиться стратификация. В качестве фактов могут выступать время, операторы, оборудование, условия производственных операций (такие как температура, влажность, давление, освещенность и т.п.), материалы и средства измерения (такие как измерительное оборудование и методы измерения).

– Определяется число страт (слоев). Количество страт берется соответственно количеству факторов, выявленных на предыдущем шаге. Например, отклонения в показателях продукции могут возникать из-за действий оператора. Если к производству продукта привлечено четыре оператора, то стратификация выполняется по четырем факторам и число страт должно быть четыре. Или, если условия производства продукта остаются одними и теми же, изменения в характеристиках могут возникать в разные периоды времени – первая смена, вторая смена или третья смена работы. В этом варианте страт будет три (по количеству смен) и стратификация проводится по трем факторам.

– Выбирается необходимый инструмент качества для графического представления статистических данных. Как правило, для этих целей используется диаграмма разброса, контрольная карта или гистограмма. Можно применять и табличный метод, но графический способ является более наглядным и позволяет быстрее определить системность в представленных данных.

– Определяется количество статистических данных, попадающих в каждую страту. Для того чтобы стратификация данных была эффективной необходимо придерживаться двух условий. Во-первых, различия между значениями случайной величины внутри страты должны быть как можно меньше по сравнению с различием ее значений в исходной совокупности данных. Во-вторых, различия между стратами должны быть как можно больше. Количественно это различие можно определить по разнице средних значений случайной величины в каждой страте.

На выбранный графический инструмент качества «наносятся» данные с указанием принадлежности этих данных к каждой из страт. Для отделения данных друг от друга, можно использовать самый простой метод – цветовую индикацию данных.

Проводится анализ подмножества данных. Анализ данных проводится для каждой страты отдельно.

Стратификацию данных рассмотрим на примере данных о браке, допущенном тремя рабочими по результатам работы каждого на трех станках в разные смены (Таблица 25).

Таблица 25

Рабочий	Станок	1 смена	2 смена	3 смена	Число дефектов на станках	Сумма дефектов рабочего
А	**А 1**	**0**	**0**	**1**	**1**	**24**
	А 2	**2**	**1**	**0**	**3**	
	А 3	**1**	**7**	**12**	**20**	
Б	**Б 1**	**2**	**7**	**6**	**15**	**45**
	Б 2	**1**	**1**	**3**	**5**	
	Б 3	**2**	**8**	**15**	**25**	
В	**В 1**	**1**	**1**	**2**	**4**	**52**
	В 2	**2**	**2**	**4**	**8**	
	В 3	**3**	**17**	**20**	**40**	

Стратификация данных в Statistica 6.0

Произведите стратификацию данных по станкам в программе Statistica 6.0. Для этого создайте таблицу распределения данных о браке по станкам. (Таблица 26).

Таблица 26

Станок 1	Станок 2	Станок 3
0	2	1
2	1	2
1	2	3
0	1	7
7	1	8
1	2	17
1	0	12
6	3	15
2	4	20

Выберитев меню Графики-Гисиограммы-2DHistograms. Задайте переменные, в окне **Тип графика** выберите значение **Multiple**, задайте число интервалов. Нажмите **OK** (рис. 106).

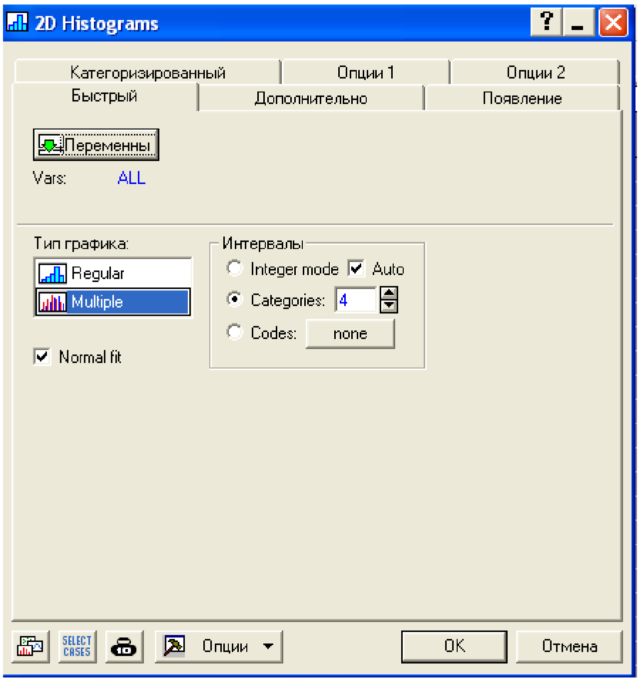


Рис. 106

В результате произведенных действий будет построена гистограмма распределения брака по станкам (рис. 107).

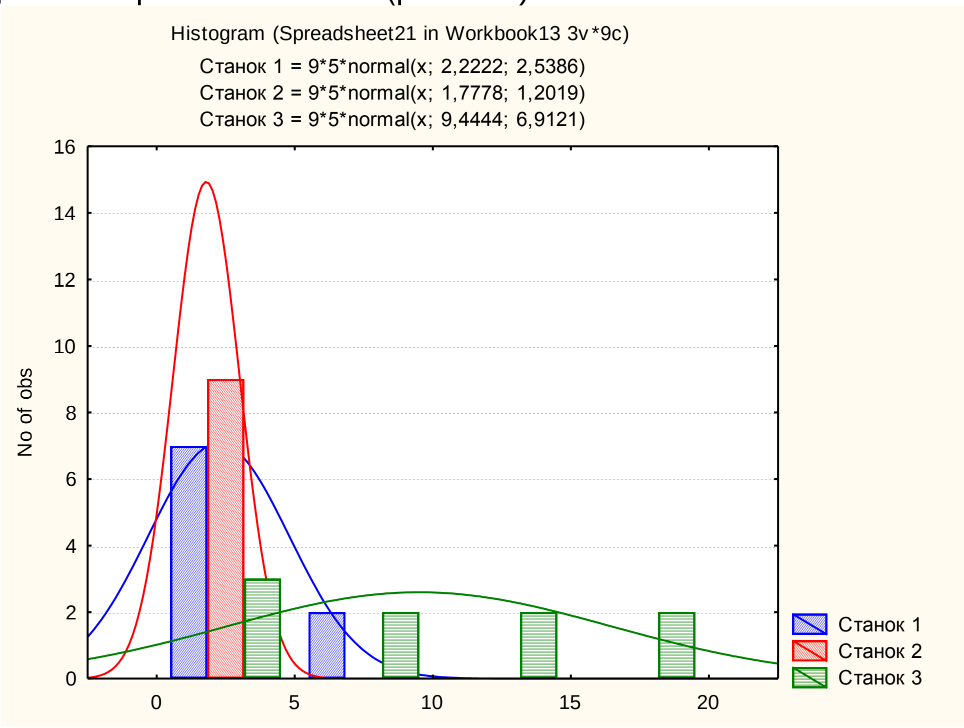


Рис. 107

<u>Задание</u>

Постройте гистограмму распределения брака по рабочим. Постройте гистограмму распределения брака по сменам. Для этого выберите соответствующие переменные.

ДИСПЕРСИОННЫЙ АНАЛИЗ

Дисперсионный анализ (от латинского Dispersio – рассеивание) – статистический метод, позволяющий анализировать влияние различных факторов на исследуемую переменную. Метод был разработан биологом Р. Фишером в 1925 году и применялся первоначально для оценки экспериментов в растениеводстве. В дальнейшем выяснилась общенаучная значимость дисперсионного анализа для экспериментов в психологии, педагогике, медицине и др.

Целью дисперсионного анализа является проверка значимости различия между средними с помощью сравнения дисперсий.

Дисперсию измеряемого признака разлагают на независимые слагаемые, каждое из которых характеризует влияние того или иного фактора или их взаимодействия. Последующее сравнение таких слагаемых позволяет оценить значимость каждого изучаемого фактора, а также их комбинации.

При истинности нулевой гипотезы (о равенстве средних в нескольких группах наблюдений, выбранных из генеральной совокупности), оценка дисперсии, связанной с внутригрупповой изменчивостью, должна быть близкой к оценке межгрупповой дисперсии.

На практике часто возникают задачи проверки существенности различий выборочных средних нескольких совокупностей. Например, требуется оценить влияние различного сырья на качество производимой продукции, решить задачу о влиянии количества удобрений на урожайность с/х продукции.

В зависимости от количества факторов, определяющих вариацию результативного признака, дисперсионный анализ подразделяют на однофакторный и многофакторный.

При проведении дисперсионного анализа должны выполняться следующие статистические допущения: **независимо от уровня фактора величины отклика имеют нормальный (Гауссовский) закон распределения и одинаковую дисперсию. При неизвестном законе распределения величин отклика используют непараметрические (чаще всего ранговые) методы анализа.**

Однофакторный дисперсионный анализ

В основе дисперсионного анализа лежит разделение дисперсии на части или компоненты. Вариацию, обусловленную влиянием фактора, положенного в основу группировки, характеризует межгрупповая дисперсия σ^2. Она является мерой вариации частных средних по группам $\bar{x}_j$ вокруг общей средней $\bar{x}$ и определяется по формуле:

$$\bar{\sigma}^2 = \frac{\sum_{j=1}^{k} (\bar{x}_j - \bar{x})^2 \times n_j}{\sum_{j=1}^{k} n_j}, \qquad (66)$$

где k - число групп;

n_j - число единиц в j-ой группе;

$\bar{x}_j$ - частная средняя по j-ой группе;

$\bar{x}$ - общая средняя по совокупности единиц.

Вариацию, обусловленную влиянием прочих факторов, характеризует в каждой группе внутригрупповая дисперсия σ_j^2.

$$\sigma_j^2 = \frac{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}{n} \tag{67}$$

Между общей дисперсией σ_0^2, внутригрупповой дисперсией σ^2 и межгрупповой дисперсией $\bar{\sigma}^2$ существует соотношение:

$\sigma_0^2 = \bar{\sigma}^2 + \sigma^2$.

Внутригрупповая дисперсия объясняет влияние неучтенных при группировке факторов, а межгрупповая дисперсия объясняет влияние факторов группировки на среднее значение по группе.

Базовая таблица 27 для проведения однофакторного дисперсионного анализа.

Таблица 27

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Средний квадрат
Межгрупповая	$Q_1 = n\sum_{i=1}^{m}(\bar{x}_{i*} - \bar{x}_{**})^2$	m-1	$= Q_1/(m-1)$
Внутригрупповая	$Q_2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i*})^2$	n-m	$= Q_2/(n-m)$
Общая	$Q = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{**})^2$	n-1	

Процедура однофакторного дисперсионного анализа состоит в проверке гипотезы H_0о том, что имеется одна группа однородных экспериментальных данных против альтернативы о том, что таких групп больше, чем одна. Под однородностью понимается одинаковость средних значений и дисперсий в любом подмножестве данных. При этом дисперсии могут быть как известны, так и неизвестны заранее.

Проверка нулевой гипотезы H_0 сводится к проверке существенности различия несмещенных выборочных оценок S_1^2 и S_2^2 дисперсии σ^2.

Для этого вычисляется статистика $F=S_1^2/S_2^2$ и сравнивается с табличным значением F- распределения Фишера для заданного уровня значимости - $F_{\alpha:K1:K2}$.

Гипотеза H_0 отвергается, если фактически вычисленное значение статистики больше критического $F_{\alpha:K1:K2}$, определенного на уровне значимости α при числе степеней свободы k_1=m-1 и k_2=n-m, и принимается, если $F<F_{\alpha:K1:K2}$.

F- распределение Фишера (для x > 0) имеет следующую функцию плотности (для ν = 1, 2, ...; ω = 1, 2, ...):

$$f(x)=\frac{\Gamma[(\nu+\omega)/2]}{\Gamma(\nu/2)*\Gamma(\omega/2)}*(\nu/\omega)^{\nu/2}*x^{(\nu/2)-1}*[1+(\nu/\omega)*x]^{-(\nu+\omega)/2}, \qquad (68)$$

где ν,ω - степени свободы;

Γ - гамма-функция.

Табличное значение критерия Фишера находят следующим образом:

Определяют k1, равное количеству факторов r.

Например, в однофакторной модели (модели парной регрессии) k1=1, в двухфакторной k=2.

Определяют k2, которое определяется по формуле n - r - 1, где n - число наблюдений, r - количество факторов. Например, в однофакторной модели k2 = n - 2.

На пересечении столбца k1 и строки k2 находят значение критерия Фишера.

Многофакторный дисперсионный анализ

Принципиальной разницы между многофакторным и однофакторным дисперсионным анализом нет. Многофакторный анализ не меняет общую логику дисперсионного анализа, а лишь несколько усложняет ее, поскольку, кроме учета влияния на зависимую переменную каждого из факторов по отдельности, следует оценивать и их совместное действие. Таким образом, то новое, что вносит в анализ данных многофакторный дисперсионный анализ, касается в основном возможности оценить межфакторное взаимодействие. Тем не менее, по-прежнему остается возможность оценивать влияние каждого фактора в отдельности. В этом смысле процедура многофакторного дисперсионного анализа (в варианте ее компьютерного использования), несомненно, более экономична, поскольку всего за один запуск решает сразу две задачи: оценивается влияние каждого из факторов и их взаимодействие.

Общая схема двухфакторного эксперимента, данные которого обрабатываются дисперсионным анализом, имеет вид (рис. 108):

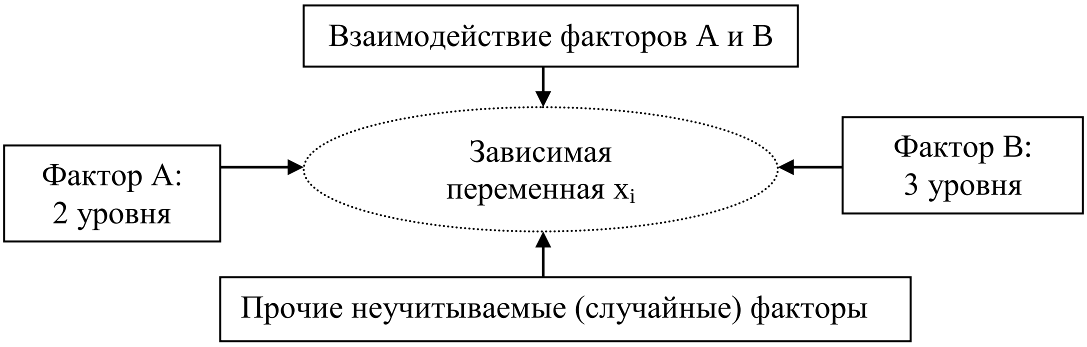


Рис. 108

Данные, подвергаемые многофакторному дисперсионному анализу, часто обозначают в соответствии с количеством факторов и их уровней.

Предположив, что в рассматриваемой задаче о качестве различных m партий изделия изготавливались на разных t станках и требуется выяснить, имеются ли существенные различия в качестве изделий по каждому фактору:

А - партия изделий;

В - станок.

В результате получается переход к задаче двухфакторного дисперсионного анализа.

Все данные представлены в таблице, в которой по строкам - уровни A_i фактора А, по столбцам — уровни B_j фактора В, а в соответствующих ячейках, таблицы 28 находятся значения показателя качества изделий x_{ijk}(i=1,2,...,m; j=1,2,...,l; k=1,2,...,n).

Таблица 28

	B_1	B_2	…	B_j	…	B_l
A_1	$x_{11l},\dots,x_{11k}$	$x_{12l},\dots,x_{12k}$	…	$x_{1jl},\dots,x_{1jk}$	…	$x_{1ll},\dots,x_{1lk}$
A_2	$x_{21l},\dots,x_{21k}$	$x_{22l},\dots,x_{22k}$	…	$x_{2jl},\dots,x_{2jk}$	…	$x_{2ll},\dots,x_{2lk}$
…	…	…	…	…	…	…
A_i	$x_{i1l},\dots,x_{i1k}$	$x_{i2l},\dots,x_{i2k}$	…	$x_{ijl},\dots,x_{ijk}$	…	$x_{jll},\dots,x_{jlk}$
…	…	…	…	…	…	…
A_m	$x_{m1l},\dots,x_{m1k}$	$x_{m2l},\dots,x_{m2k}$	…	$x_{mjl},\dots,x_{mjk}$	…	$x_{mll},\dots,x_{mlk}$

Двухфакторная дисперсионная модель имеет вид:

$$x_{ijk}=\mu+F_i+G_j+I_{ij}+\varepsilon_{ijk}, \qquad (69)$$

где x_{ijk} - значение наблюдения в ячейке ij с номером k;

μ - общая средняя;

F_i- эффект, обусловленный влиянием i-го уровня фактора A;

G_j - эффект, обусловленный влиянием j-го уровня фактора B;
I_{ij} - эффект, обусловленный взаимодействием двух факторов, т.е. отклонение от средней по наблюдениям в ячейке ij от суммы первых трех слагаемых в модели (15);
ε_{ijk} - возмущение, обусловленное вариацией переменной внутри отдельной ячейки.

Предполагается, что ε_{ijk} имеет нормальный закон распределения N(0; c^2), а все математические ожидания F_*,G_*, I_{i*}, I_{*j} равны нулю.

Групповые средние находятся по формулам:

- в ячейке:

$$\bar{x}_{ij*}=\frac{\sum_{k=1}^{n}x_{ijk}}{n}, \qquad (70)$$

по строке:

$$\bar{x}_{i**}=\frac{\sum_{j=1}^{l}\bar{x}_{ij*}}{l}, \qquad (71)$$

по столбцу:

$$\bar{x}_{*j*}=\frac{\sum_{i=1}^{m}\bar{x}_{ij*}}{m}, \qquad (72)$$

общая средняя:

$$\bar{x}_{***}=\frac{\sum_{i=1}^{m}\sum_{j=1}^{l}\bar{x}_{ij*}}{ml}. \qquad (73)$$

В таблице 29 представлен общий вид вычисления значений, с помощью дисперсионного анализа.

Таблица 29

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Средние квадраты
Межгрупповая (фактор А)	$Q_1 = \ln\sum_{i=1}^{m}(\bar{x}_{i**} - \bar{x}_{***})^2$	m-1	$S_1^2 = \frac{Q_1}{m-1}$
Межгрупповая (фактор В)	$Q_2 = mn\sum_{j=1}^{l}(\bar{x}_{*j*} - \bar{x}_{***})^2$	l-1	$S_2^2 = \frac{Q_2}{l-1}$
Взаимодействие	$Q_3 = n\sum_{i=1}^{m}\sum_{j=1}^{l}(\bar{x}_{ij*} - \bar{x}_{i**} - \bar{x}_{*j*} + \bar{x}_{***})^2$	(m-1)(l-1)	$S_3^2 = \frac{Q_3}{(m-1)(l-1)}$
Остаточная	$Q_4 = \sum_{i=1}^{m}\sum_{j=1}^{l}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{ij*})^2$	mln - ml	$S_4^2 = \frac{Q_4}{m\ln - ml}$
Общая	$Q = \sum_{i=1}^{m}\sum_{j=1}^{l}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{***})^2$	mln - 1	

Проверка нулевых гипотез НА, НВ, НАВ об отсутствии влияния на рассматриваемую переменную факторов А, В и их взаимодействия АВ осуществляется сравнением отношений $\frac{S_1^2}{S_4^2}$, $\frac{S_2^2}{S_4^2}$, $\frac{S_3^2}{S_4^2}$ (для модели с фиксированными уровнями факторов) или отношений $\frac{S_1^2}{S_3^2}$, $\frac{S_2^2}{S_3^2}$, $\frac{S_3^2}{S_4^2}$ (для случайной модели) с соответствующими табличными значениями F – критерия Фишера – Снедекора.

Отклонение от основных предпосылок дисперсионного анализа — нормальности распределения исследуемой переменной и равенства дисперсий в ячейках (если оно не чрезмерное) — не сказывается существенно на результатах дисперсионного анализа при равном числе наблюдений в ячейках, но может быть очень чувствительно при неравном их числе.

Проведение дисперсионного анализа в MSEXCEL

По алгоритму базовой таблицы проведите однофакторный дисперсионный анализ по данным о браке на станках. (Таблица 30)

Таблица 30

Станок 1 x	Станок 2 y	Станок 3 z	$(x-\bar{x})$^2	$(y-\bar{y})$^2	$(z-\bar{z})$^2
0	2	1	4,94	0,05	71,31
2	1	2	0,05	0,60	55,42
1	2	3	1,49	0,05	41,53
0	1	7	4,94	0,60	5,98
7	1	8	22,83	0,60	2,09
1	2	17	1,49	0,05	57,09
1	0	12	1,49	3,16	6,53
6	3	15	14,27	1,49	30,86
2	4	20	0,05	4,94	111,42
Среднее значение	Среднее значение	Среднее значение	Сумма	Сумма	Сумма
2,22	1,78	9,44	51,56	11,56	382,22
Внутригрупповая дисперсия Q_2			Сумма		
			445,33		
Общее среднее $\bar{\bar{a}}$			4,48		
$(x-\bar{\bar{a}})$^2	$(y-\bar{\bar{a}})$^2	$(z-\bar{\bar{a}})$^2	Сумма		
5,10	7,31	24,63	37,05		
Количество значений n			9		
Межгрупповая дисперсия Q_1			333,407		
Число групп m			3		
S_1^2 =Q$_1$/(m-1)			166,70		
S_2^2 = Q$_2$/(n-m)			18,56		
$F=\dfrac{S_1^2}{S_2^2}$			8,98		

По таблице значений F-критерия Фишера (Приложение 4) при уровне значимости $\alpha = 0{,}05$ определите F критическое =3,403.

Поскольку расчетное значение статистики

нулевая гипотеза H_0 об отсутствии различия однородности групп отвергается.

Произведенный расчет достаточно трудоемкий. Произведите однофакторный дисперсионный анализ в MS Excel с использованием надстройки Анализ данных. Для этого задайте входной интервал, выберите способ группирования, значение и выходной интервал (рис. 109).

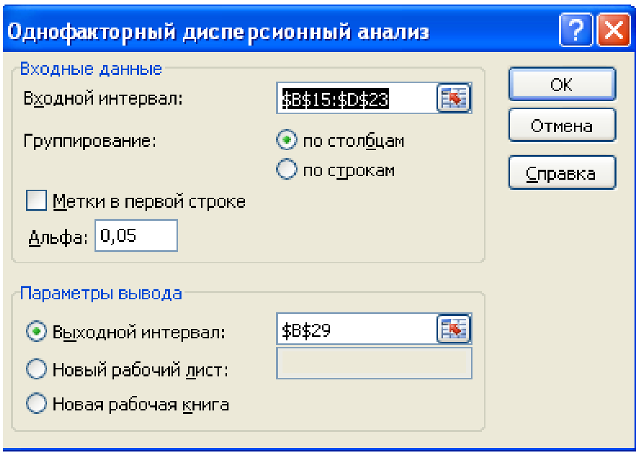


Рис. 109

Сравните полученные результаты с результатами произведенных расчетов

Однофакторный дисперсионный анализ станки

ИТОГИ

Группы	*Счет*	*Сумма*	*Среднее*	*Дисперсия*
Столбец 1	9	20	2,2222	6,4444
Столбец 2	9	16	1,7778	1,4444
Столбец 3	9	85	9,4444	47,7778

Дисперсионный анализ

Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Между группами	333,4074	2	166,7037	8,9840	0,0012	3,4028
Внутри групп	445,3333	24	18,5556			
Итого	778,7407	26				

Для проведения двухфакторного дисперсионного анализа без повторений рабочие – станки составим таблицу 31 данных

Таблица 31

	Рабочий 1	Рабочий 2	Рабочий 3
Станок 1	1	15	4
Станок 2	3	5	8
Станок 3	20	25	40

ИТОГИ	*Счет*	*Сумма*	*Среднее*	*Дисперсия*
Строка 1	3	20	6,6667	54,3333
Строка 2	3	16	5,3333	6,3333
Строка 3	3	85	28,3333	108,3333
Столбец 1	3	24	8	109
Столбец 2	3	45	15	100
Столбец 3	3	52	17,333333	389,333

Дисперсионный анализ

Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Строки	1000,2222	2	500,1111	10,1833	0,0269	6,9443
Столбцы	141,5556	2	70,7778	1,4412	0,3378	6,9443
Погрешность	196,4444	4	49,1111			
Итого	1338,2222	8				

Для проведения двухфакторного дисперсионного анализа с повторениями Смена-Станок составьте таблицу 32 исходных данных

Таблица 32

	Смена 1	Смена 2	Смена 3
Станок 1	**0**	**0**	**1**
Станок 2	**2**	**1**	**0**
Станок 3	**1**	**7**	**12**
Станок 1	**2**	**7**	**6**
Станок 2	**1**	**1**	**3**
Станок 3	**2**	**8**	**15**
Станок 1	**1**	**1**	**2**
Станок 2	**2**	**2**	**4**
Станок 3	**3**	**17**	**20**

Задайте входной интервал, число строк для выборки - 3, значение и выходной интервал.

ИТОГИ	Смена 1	Смена 2	Смена 3	Итого
Станок 1				
Счет	3	3	3	9
Сумма	3	8	13	24
Среднее	1	2,6667	4,3333	2,6667
Дисперсия	1	14,3333	44,3333	17
Станок 1				
Счет	3	3	3	9
Сумма	5	16	24	45
Среднее	1,66667	5,3333	8	5
Дисперсия	0,33333	14,333	39	21
Станок 1				
Счет	3	3	3	9
Сумма	6	20	26	52
Среднее	2	6,6667	8,6667	5,7778
Дисперсия	1	80,3333	97,3333	53,4444
Итого				
Счет	9	9	9	
Сумма	14	44	63	
Среднее	1,55556	4,8889	7	
Дисперсия	0,77778	30,361	49,25	

Дисперсионный анализ

Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Выборка	47,1852	2	23,5926	0,7272	0,4969	3,5546
Столбцы	135,6296	2	67,8148	2,0902	0,1527	3,5546
Взаимодействие	11,9259	4	2,9815	0,0919	0,9838	2,9277
Внутри	584	18	32,4444			
Итого	778,7407	26				

Однофакторный дисперсионный анализ в Statistica 6.0

Введите таблицу исходных данных Станок-Брак для проведения однофакторного дисперсионного анализа в Statistica 6.0 (рис. 110).

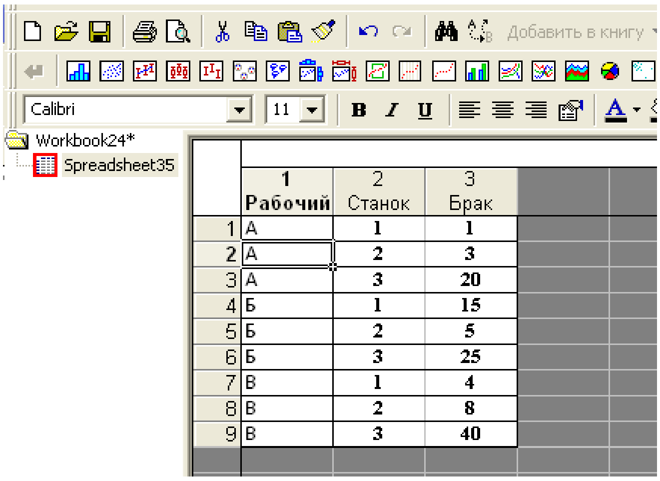


Рис. 110

Выберите в меню **Статистика-Основная статистика/Таблицы-Breakdown&one-way ANOVA**. Нажмите **ОК** (рис. 111).

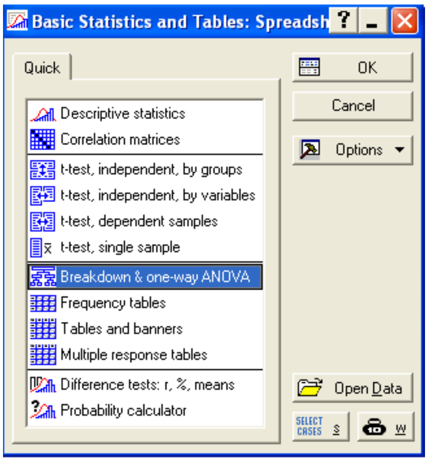


Рис. 111

Выберите для анализа зависимую и группирующую переменную. Нажмите **ОК** (рис. 112).

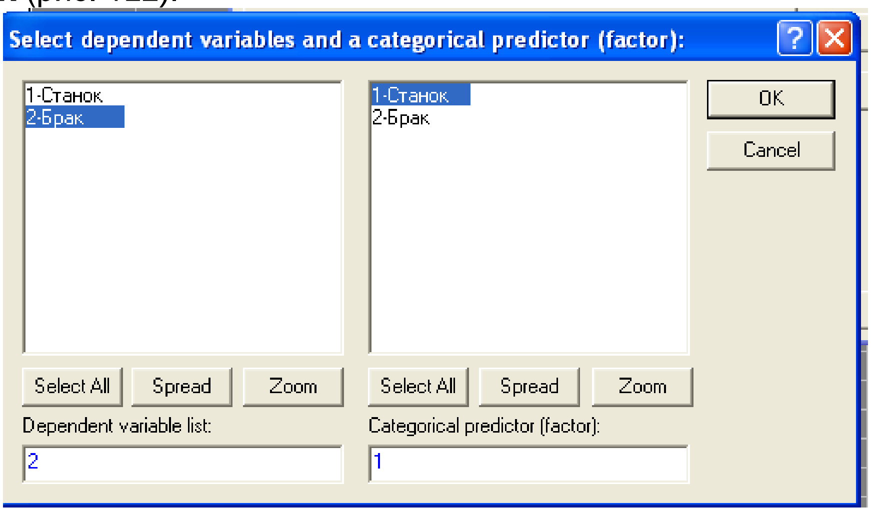


Рис. 112

Выберите список значений независимой переменной. Для этого нажмите на кнопку **All.** Нажмите **ОК** (рис. 113).

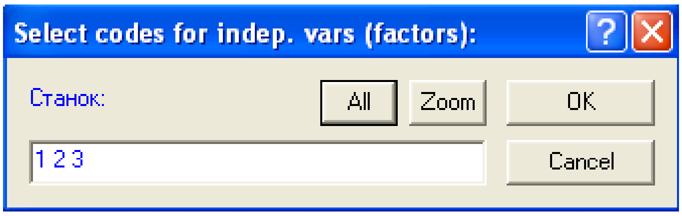


Рис. 113

После выбора всех параметров нажмите **ОК.** В открывшемся окне перейдите на вкладку ANOVA&tests (рис. 114).

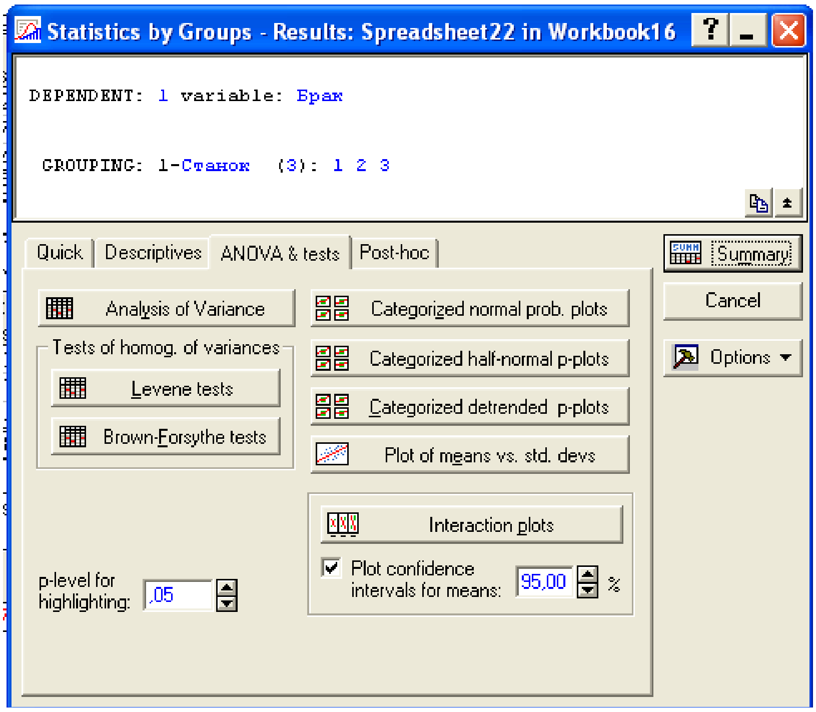


Рис. 114

Нажмите на кнопку Analysis of Variance (рис. 115).

	Analysis of Variance (Spreadsheet35 in Workbook24) Marked effects are significant at p < ,05000							
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
Брак	1000,22	2	500,111	338,000	6	56,3333	8,87771	0,01611

Рис. 115

Многофакторный дисперсионный анализ в программе Statistica 6.0

Выберите в меню **Статистика – Анализ вариантов- Main effects ANOVA**. Нажмите **ОК** (рис. 116).

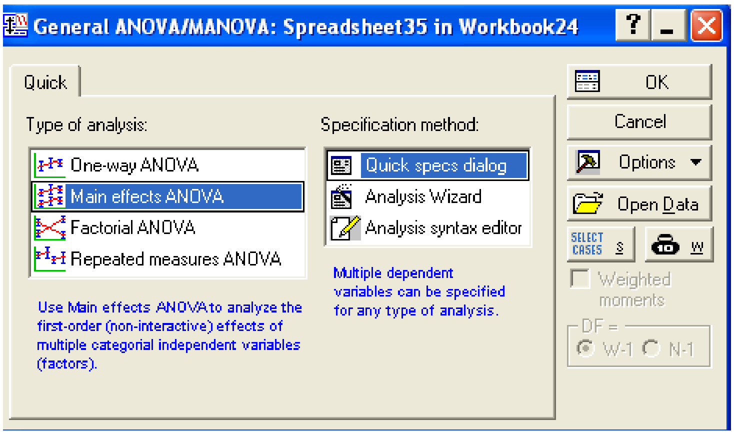


Рис. 116

Выберите переменные для анализа. Нажмите **ОК** (рис. 117).

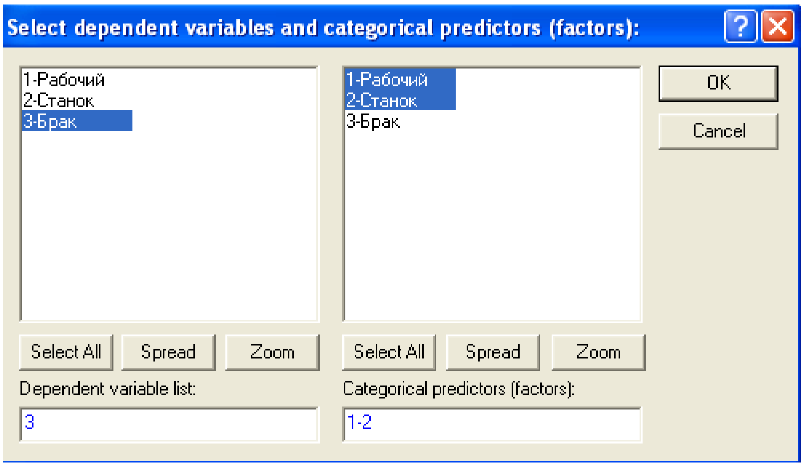


Рис. 117

Выберите список значений независимых переменных. Для этого нажмите на кнопки **All.** Нажмите **ОК** (рис. 118).

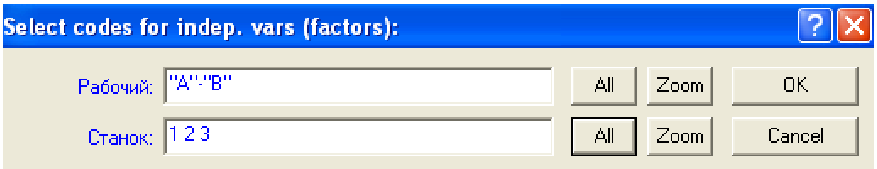


Рис. 118

Результат анализа представлен на рисунке 119.

	Univariate Tests of Significance for Брак (Spreads Sigma-restricted parameterization Effective hypothesis decomposition				
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	1626,778	1	1626,778	33,12443	0,004520
Рабочий	141,556	2	70,778	1,44118	0,337789
Станок	1000,222	2	500,111	10,18326	0,026948
Error	196,444	4	49,111		

Рис. 119

НЕПАРАМЕТРИЧЕСКИЕ СТАТИСТИЧЕСКИЕ КРИТЕРИИ (NONPARAMETRIC STATISTICAL TESTS).

По сравнению со стандартными параметрическими процедурами, непараметрические статистические методы основываются на более слабых допущениях в отношении анализируемых данных. Существует определенное соотношение выгод и потерь, связанных с использованием непараметрических статистических критериев вместо параметрических. Главным мотивом применения непараметрических методов служит нежелание делать допущения, необходимые для использования параметрических процедур. Дополнительным соображением в пользу выбора непараметрических статистических критериев для части исследователей служит присущая некоторым таким критериям легкость применения и простота вычислений. Однако с использованием непараметрических критериев связаны определенные неудобства и потери. Прежде всего, проверяемая с помощью непараметрического критерия нулевая гипотеза обычно не является в точности той же самой нулевой гипотезой, которая проверяется при использовании соответствующего параметрического критерия. Нулевая гипотеза при применении t-критерия для независимых выборок формулируется следующим образом: средние двух генеральных совокупностей равны. Нулевая гипотеза при использовании медианного критерия или критерия Манна-Уитни, которые можно было бы применить к тем же данным для определения того, будут ли две группы оценок "значимо различаться между собой", звучит иначе: две генеральные совокупности тождественны. А это предполагает, что выявление значимого различия могло оказаться следствием какого-то неизвестного нам сочетания различий в центральной тенденции, вариабельности и симметрии. Кроме того, непараметрические критерии могут быть нечувствительными к некоторым видам различий между совокупностями. Другое слабое место непараметрических критериев заключается в их относительно низкой статистической мощности по сравнению со стандартными параметрическими критериями. Мощность статистического критерия определяется как вероятность отклонения нулевой гипотезы в тех случаях, когда она является ложной. Непараметрические критерии обычно требуют больших объемов выборки, чтобы сравняться по статистической мощности с параметрическими критериями. Когда анализируемые данные более или менее соответствуют допущениям параметрических критериев, следует, по всей вероятности, использовать именно эти критерии. Простых рецептов в отношении того, в каких ситуациях следует применять именно непараметрические статистические критерии, не существует. Чтобы сделать оптимальный выбор в конкретной ситуации, исследователь должен знать характеристики анализируемых данных и располагать информацией о доступных параметрических и непараметрических критериях.

Примеры непараметрических критериев

В большинстве непараметрических статистических критериев исходные оценки или результаты наблюдения заменяются другой переменной, содержащей меньше информации. Один важный класс непараметрических методов составляют критерии, использующие порядковые свойства данных. Другой важный класс образуют критерии, использующие только информацию о том, будет ли результат наблюдения выше или ниже некой фиксированной величины, скажем, медианы. Еще один класс критериев основан на частоте появления "серий" ("runs") в совокупности данных. Серия - ряд событий одного типа, появляющихся подряд как часть упорядоченной последовательности событий. Упорядочение может быть временным или основываться на величине оценок. Исследование серий может быть полезным при решении вопроса о случайном или неслучайном характере последовательности наблюдений. Хотя существует огромное множество непараметрических критериев, которые можно применять в самых разных ситуациях, лишь немногие из часто встречающихся критериев допускают краткое описание.

Критерий знаков для сравнения групп с попарно связанными вариантами

В отличие от параметрического t-критерия для сопряженных пар, который предполагает использование фактических значений разностей элементов каждой пары, критерий знаков учитывает только знак (+ или -) этих разностей. Цель - определить, есть ли преобладание любого из знаков, а проверяемая нулевая гипотеза состоит в том, что вероятность появления "плюса" равна вероятности появления "минуса".

Медианные критерии для независимых групп и групп с попарно связанными вариантами

Медианные критерии предполагают сравнение нескольких выборок на основе отклонений от медианы. Проверяемая нулевая гипотеза состоит в том, что разные генеральные совокупности, из которых извлекаются сравниваемые выборки, являются идентичными. Существуют формы медианного критерия для независимых групп и групп с попарно связанными вариантами.

Т - критерий Вилкоксона для групп с попарно связанными вариантами

Т - критерий Вилкоксона основан на использовании рангов абсолютных значений разностей между членами сопряженных пар. Этот критерий является хорошей альтернативой t-критерию для коррелированных величин.

Критерий Фридмана

Критерий Фридмана для групп с попарно связанными вариантами Критерий Фридмана можно рассматривать как обобщение критерия Уилкокс для сравнения нескольких (>2) групп. Этот критерий представляет собой хорошую альтернативу параметрическому дисперсионному анализу с повторными наблюдениями.

Вычисление. Для применения этого критерия столбцы таблицы данных отражают различные значения переменной эффекта, а строки соответствуют повторным измерениям одного и того же субъекта. С помощью критерия Фридмана мы проверяем нулевую гипотезу о том, что различные методы лечения дают практически одинаковые результаты. Процедура состоит в упорядочивании (ранжировании) значений в каждой строке (при этом ранги в каждой строке принимают значения от 1 до m - число сравниваемых методов лечения), суммировании полученных рангов по каждому столбцу и вычислении статистики Хи-квадрат. Рассчитанная статистика Хи-квадрат имеет такое же распределение, что и Хи-квадрат при (m-1) степенях свободы. Если соответствующее значение превзойдет критическое значение (для выбранного уровня значимости и соответствующего числа степеней свобод), то нулевая гипотеза отклоняется.

Критерий Манна-Уитни для двух независимых групп

Критерий Манна - Уитни основан на использовании рангов результатов наблюдений с целью проверки гипотез в отношении двух генеральных совокупностей, из которых извлекаются независимые выборки сравниваемых наблюдений.

Ранговый дисперсионный анализ по Краскелу-Уоллесу

Этот критерий можно рассматривать как обобщение критерия Манна - Уитни для сравнения неск. (>2) независимых выборок. Нулевая гипотеза формулируется следующим образом: k независимых выборок объема n1 , п2 , ..., nk являются выборками из идентичных генеральных совокупностей.

U-критерий Манна-Уитни: две зависимые выборки

Настоящий статистический метод был предложен Фрэнком Вилкоксоном в 1945 году. Однако в 1947 году метод был улучшен и расширен Х. Б. Манном и Д. Р. Уитни, посему U-критерий чаще называют их именами.

Критерий предназначен для оценки различий между двумя выборками по уровню какого-либо признака, количественно измеренного. Он позволяет выявлять различия между малыми выборками, когда n1,n2≥3 или n1=2, n2≥5.

Существует несколько способов использования критерия и несколько вариантов таблиц критических значений, соответствующих этим способам (Гублер Е. В., 1978; Рунион Р., 1982; Захаров В. П., 1985; McCall R., 1970; Krauth J., 1988).

Этот метод определяет, достаточно ли мала зона перекрещивающихся значений между двумя рядами.

1-м рядом (выборкой, группой) мы называем тот ряд значений, в котором значения, по предварительной оценке, выше, а 2-м рядом - тот, где они предположительно ниже.

Чем меньше область перекрещивающихся значений, тем более вероятно, что различия достоверны. Иногда эти различия называют различиями в расположении двух выборок (Welkowitz J. et al., 1982).

Эмпирическое значение критерия U отражает то, насколько велика зона совпадения между рядами. Поэтому чем меньше $U_{эмп}$, тем более вероятно, что различия достоверны.

В каждой из выборок должно быть не менее 3 значений признака. Допускается, чтобы в одной выборке было два значения, но во второй тогда не менее пяти.

В выборочных данных не должно быть совпадающих значений (все числа — разные) или таких совпадений должно быть очень мало.

Гипотезы U - кН0: Уровень признака в группе 2 не ниже уровня признака в группе 1.

Н1: **Уровень признака в группе 2 ниже уровня признака в группе 1.**

Для применения U-критерия Манна — Уитни нужно произвести следующие операции.

1) Составить единый ранжированный ряд из обеих сопоставляемых выборок, расставив их элементы по степени нарастания признака и приписав меньшему значению меньший ранг. Общее количество рангов получится равным:

$$N = n1 + n2 \qquad (74)$$

где n1— количество единиц в первой выборке,
n2 — количество единиц во второй выборке.

2) Разделить единый ранжированный ряд на два, состоящие соответственно из единиц первой и второй выборок. Подсчитать отдельно сумму рангов, пришедшихся на долю элементов первой выборки, и отдельно — на долю элементов второй выборки. Определить большую из двух ранговых сумм ( ), соответствующую выборке с единиц.

3) Определить значение U-критерия Манна-Уитни по формуле:

$$U = n1 \times n2 + \frac{nx \times (nx+1)}{2} - Tx \qquad (75)$$

4) По таблице для избранного уровня статистической значимости определить критическое значение критерия для данных *n*1 и *n*2. Если полученное значение *U* **меньше** табличного или равно ему, то признается наличие существенного различия между уровнем признака в рассматриваемых выборках (принимается альтернативная гипотеза). Если же полученное значение *U* больше табличного, принимается нулевая гипотеза. Достоверность различий тем выше, чем меньше значение *U*.

5) При справедливости нулевой гипотезы критерий имеет математическое ожидание M(U) = (n1*n2)/2 и дисперсию D(U) = n1*n2*(n1+n2)/12 и при достаточно большом объёме выборочных данных (n1 > 19, n2 > 19) распределён практически нормально.

Пример подсчета U-критерия Манна-Уитни

Итак, у вас есть данные обследования, полученные в двух опытах (или в двух замерах), но на одной и той же группе испытуемых (подопытных, объектов и т.д.).

Понятие "зависимые выборки"

Две выборки считаются зависимыми друг от друга, если каждому значению одной выборки можно однозначно поставить в соответствие ровно одно значение другой выборки. Аналогично определяется зависимость нескольких выборок.

Или такое определение:

Зависимые (связанные, попарно сопряженные) выборки - это выборки, представляющие собой параметры одной и той же совокупности до и после воздействия некоторого фактора.

Чаще всего зависимые выборки – это измерения одной и той же группы объектов в разные моменты времени (например, до и после воздействия какого-либо фактора). Таким образом, зависимые выборки всегда должны содержать одинаковое количество наблюдений. В электронной таблице зависимые переменные располагаются в разных столбцах одной таблицы под разными названиями (например, показатели чего-то до воздействия и показатели чего-то после воздействия).

И вам надо из этих двух столбиков данных получить какие-то обобщённые результаты, сделать выводы. И самое главное - вам надо сравнить между собой две эти выборки. Например (Таблица 33):

Таблица 33

ФИО	Замер 1	Замер 2
Иванов	5,05	7,20
Петров	6,48	7,43
Сидоров	5,16	5,58
Николаев	7,30	7,46
Сергеев	4,70	7,05
Павлов	7,25	12,95
Семенов	5,85	5,55
Фролов	6,62	9,85
Григорьев	5,15	7,50
Пушкарев	4,83	6,38
Сазонов	6,20	14,35

1) Выделить результаты замеров ("после" - "до") - каждый своим цветом (Таблица34).

(Таблица 34)

ФИО	Замер 1	Замер 2
Иванов	5.05	7.20
Петров	6.48	7.43
Сидоров	5.16	5.58
Николаев	7.3	7.46
Сергеев	4.7	7.05
Павлов	7.25	12.95
Семенов	5.85	5.55
Фролов	6.62	9.85
Григорьев	5.15	7.50
Пушкарев	4.83	6.38
Сазонов	6.2	14.35

2) Объединить значения в один столбец, который упорядочить по возрастанию.

3) Проранжировать полученные данные, начисляя меньшему значению меньший ранг. Получится таблица рангов с числом значений равным: $N = n1 + n2 = 22$.(Таблица 35).

Таблица 35

Группа	Значения	Ранг
Замер 1	4.7	1
Замер 1	4.83	2
Замер 1	5.05	3
Замер 1	5.15	4
Замер 1	5.16	5
Замер 2	5.55	6
Замер 2	5.58	7
Замер 1	5.85	8
Замер 1	6.2	9
Замер 2	6.38	10
Замер 1	6.48	11
Замер 1	6.62	12
Замер 2	7.05	13
Замер 2	7.20	14
Замер 15	7.25	15
Замер 1	7.3	16
Замер 2	7.43	17
Замер 2	7.46	18
Замер 2	7.50	19
Замер 2	9.85	20
Замер 2	12.95	21
Замер 2	14.35	22

4) Определить ранговые суммы , Большая из рагновых сумм соответствует замеру 2.

5) Определить сумму рангов по таблице. Проверить совпадение полученной суммы рангов с расчетной.

$$Tx = Tx1 + Tx2 = 253$$

6) Определить значение U-критерия Манна-Уитни по формуле:

$$U = n1 \times n2 + \frac{nx \times (nx + 1)}{2} - Tx = 11 \times 11 + \frac{11 + (11 + 1)}{2} - 167 = 20$$

7) Определить критическое значение по таблице Критические значения критерия U Манна-Уитни для данного n по таблице для заданного уровня значимости 0,05. Ткр=30.

Поскольку Тэмп=20 меньше или равен, признается наличие существенного различия между уровнем признака в рассматриваемых выборках (принимается альтернативная гипотеза).

U-критерий Манна — Уитни в программе Statistica 6.0

Для сравнения двух способов обработки (два уровня фактора) воспользуемся ***статистикой Манна – Уитни***, реализованной в данной системе.

В критерии Манна – Уитни сформулируем нулевую гипотезу исходные две выборки – однородны, соответственно гипотеза утверждает, что выборки не однородны, т. е. влияние фактора значимо (рис. 120).

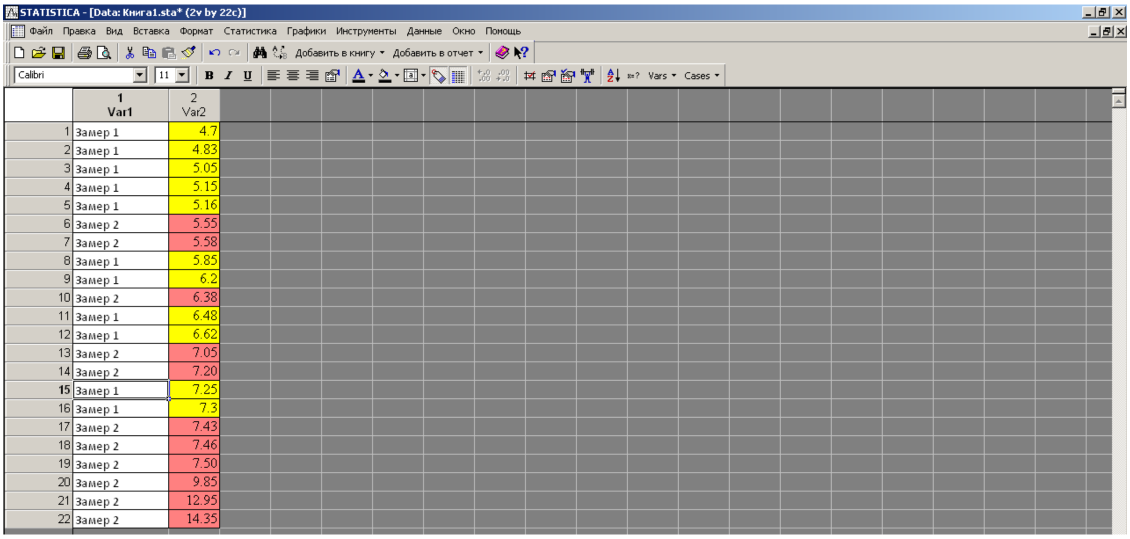


Рис. 120

Данный критерий проверяется в модуле: ***Statistics→Nonparametrics→Comparing Two Independent Samples (Groups)*** (рис. 121).

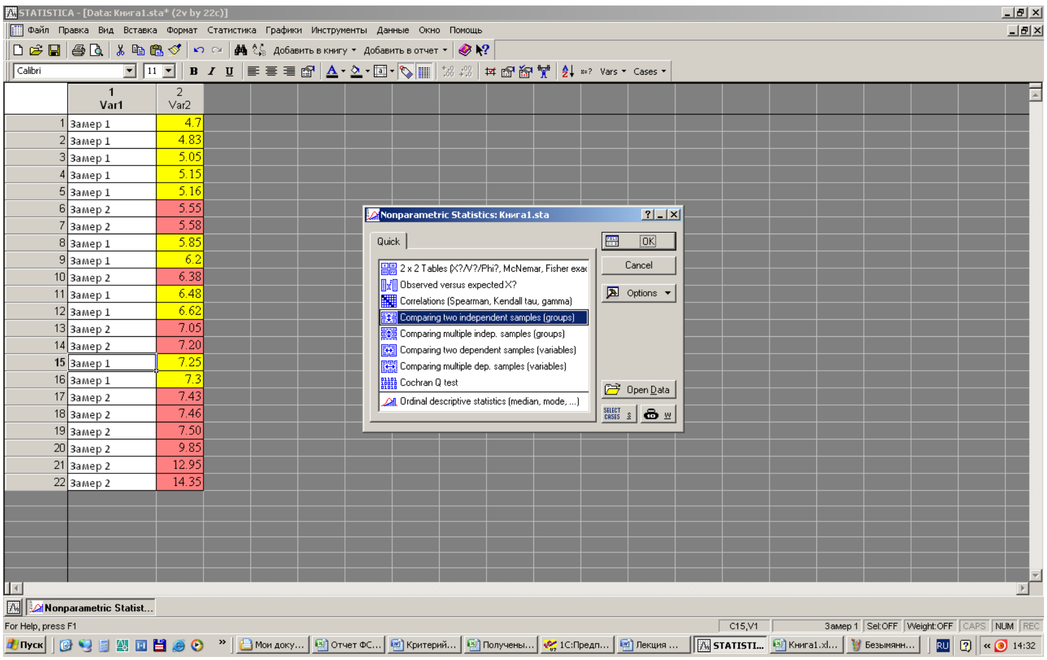


Рис. 121

С помощью клавиши **Variables** выбираем *Dependent variable* – **отклики** (*Power*) –значения столбца 2 и Indep.(grouping) variable – **уровни фактор -** значение столбца 1 (рис. 122).

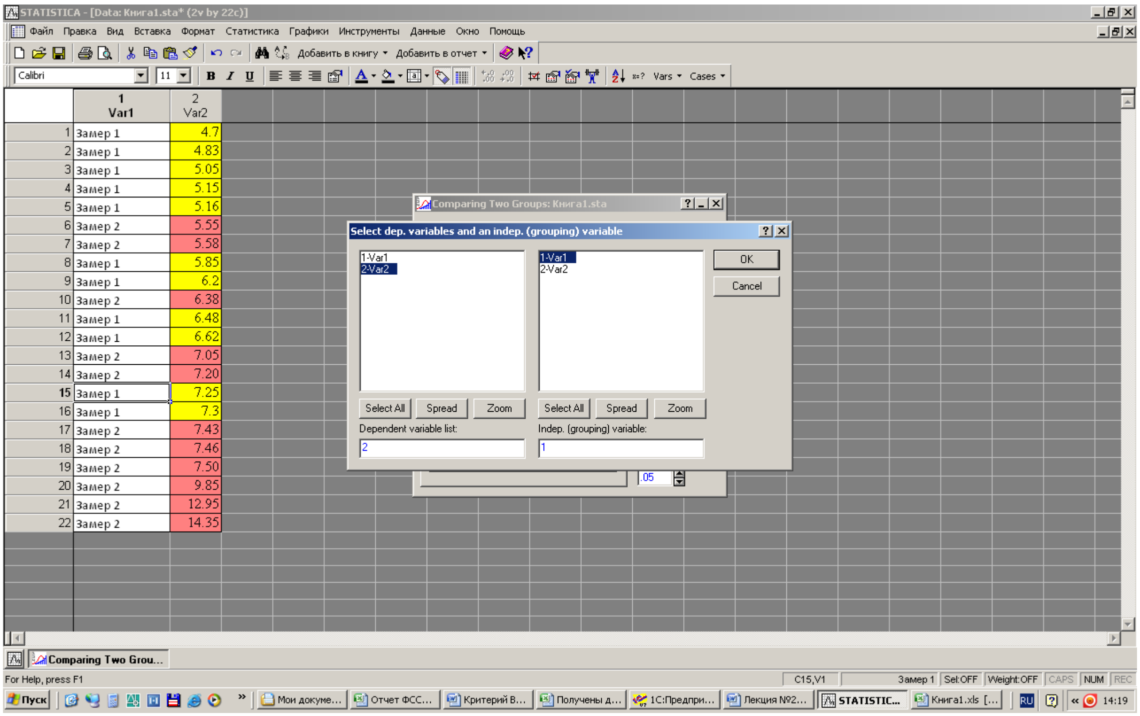


Рис. 122

После данных действий, нажмите на кнопку **Mann-Whitney U test** либо **M-W U Test** (рис. 123).

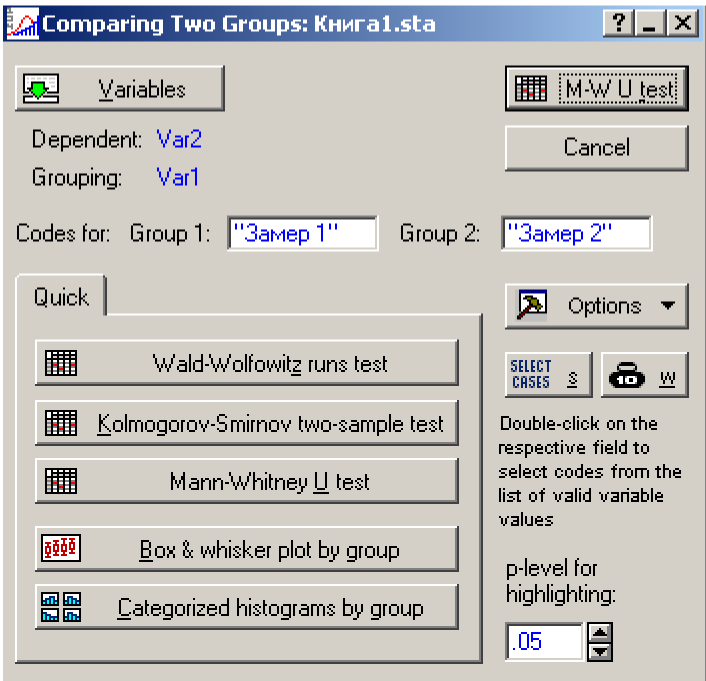


Рис. 123

Получим результат тестирования (рис. 124).

	Mann-Whitney U Test (Книга1.sta) By variable Var1 Marked tests are significant at p <.05000				
variable	Rank Sum Замер 1	Rank Sum Замер 2	U	Z	p-level
Var2	86.0000(	167.000(	20.0000(	-2.6594:	0.00782ɛ

Рис. 124

В приведенных таблицах приняты следующие обозначения:

Rank SumT$_i$– сумма рангов выборки T_i;

U–статистика Манна - Уитни для малых выборок;

p - level – вероятность принятия гипотезы H_0;

p - level – скорректированная вероятность принятия гипотезы H_0;

Rank SumT$_j$ – сумма рангов выборки T_j;

Z – нормальная аппроксимация статистики Манна - Уитни для больших выборок;

Zadjusted – скорректированная нормальная аппроксимация статистики Манна - Уитни;

Valid N – объем выборки;

2*1 sidedexactp – здесь вероятность **p** равна 1 минус кумулятивная односторонняя вероятность соответствующей статистики Манна – Уитни.

ЗАКЛЮЧЕНИЕ

В современном мире проблема качества продукции приобретает первостепенное значение. Ее успешное решение - важный показатель конкурентоспособности предприятия – производителя.

Качество продукции закладывается в процессе НИОКР, конструкторско-технологической подготовки производства, обеспечивается организацией производства.

Статистические методы контроля направлены на достижение заданного уровня качества продукции путем предупреждения причин появления несоответствий в процессе ее производства.

Рассмотренные в учебном пособии семь инструментов - простые и надежные средства решения производственных проблем обеспечения качества, рассчитаны на массовое применение, не требуют больших затрат.

Контрольный листок является инструментом сбора данных и их упорядочения для облегчения дальнейшего использования собранной информации.

Гистограмма позволяет зрительно оценить закон распределения, удобна для визуальной оценки расположения статистических данных в пределах допуска.

Контрольные карты дают возможность судить о состоянии технологического процесса по результатам выборочного контроля, контролировать значения параметров, проверять стабильность процессов, принимать корректировочные меры.

Диаграмма Парето позволяет выявить основные причины возникающих проблем и направить усилия на их разрешение.

Причинно-следственная диаграмма обеспечивает системный подход к определению фактических причин возникновения проблем.

Диаграмма разброса дает возможность определить вид и тесноту связи между парами соответствующих переменных: при наличии корреляционной зависимости между двумя факторами значительно облегчается контроль процесса.

Стратификация служит для выявления какой-либо закономерности в массиве данных за счет их разделения по факторам, применение которых зависит от конкретных задач. Например, для массива данных о возникающем в ходе производства браке, стратификация может проводиться по таким факторам как квалификация персонала, используемое оборудование, инструмент.

В учебном пособии подробно рассмотрена обработка статистических данных в программах MS Office Excel и Statistica 6.0.

MS Office Excelвключает набор инструментов статистического анализа: программную надстройку «Пакет анализа» и библиотеку статистических функций. Такого набора инструментов бывает, как правило, вполне достаточно для проведения довольно полного и качественного статистического анализа информации.

Если же пользователя не удовлетворяют подобные возможности MS Office Excel, следует обратиться к мощным специализированным пакетам, включающим большое количество методов статистического анализа, таким, как Statistica 6.0.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК ЛИТЕРАТУРЫ

1) Гмурман В. Е. Теория вероятностей и математическая статистика : учеб. пособие для вузов / В. Е. Гмурман .— 6-е изд., стер. — М. : Высш. шк., 1997 .— 479 с.

2) Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учеб. пособие для втузов / В. Е. Гмурман .— 3-е изд., перераб. и доп. — М. : Высшая школа, 1979 .— 400 с.

3) Минько А.А. Статистический анализ в MSEXCEL – М.: Издательский дом «Вильямс», 2004 – 448 с.

4) Халафян А.А. Statistica 6 Статистический анализ данных Учебник Москва Издательство «Бином» 2010 - 522 с.

5) Л. Закс. Статистическое оценивание. Пер. с нем. В.Н. Варыгина. Под ред. Ю.П. Адлера, В.Г. Горского. М., «Статистика», 1976 – 598 с.

6) Б. Хэнсен. Контроль качества. Теория и применение. Перевод с англ. М: Изд-во «Прогресс», 1968 – 519 с.

7) Э. Шиндовский, О. Шюрц. Статистические методы управления качеством. Контрольные карты и планы контроля. Пер. с нем. М: Изд-во «Мир», 1976 – 597 с.

8) ГОСТ Р 50779.21-96. Статистические методы. Правила определения и методы расчета статистических характеристик по выборочным данным. Часть 1. Нормальное распределение.

9) ГОСТ Р 50779.42-99 (ИСО 8258-91) Статистические методы. Контрольные карты Шухарта.

10) ГОСТ Р 51814.3-2001 Системы качества в автомобилестроении. Методы статистического управления процессами.

11) МС 9004-4:1993. Административное управление качеством и элементы системы качества. Часть 4. Руководящие указания по улучшению качества.

Приложение 1

ЗАДАНИЕ НА ФОРМИРОВАНИЕ ВЫБОРКИ ИНДИВИДУАЛЬНЫХ ЗНАЧЕНИЙ

№ варианта	Размер	μ	σ
1	$19h7_{-0,021}$	18,9895	0,0035
2	$19e7_{-0,061}^{-0,040}$	18,9495	0,0035
3	$32e7_{-0,075}^{-0,050}$	31,9375	0,0042
4	$32h7_{-0,050}^{-0,025}$	31,9625	0,0042
5	$12h8_{-0,027}$	11,9865	0,0045
6	$12e8_{-0,059}^{-0,032}$	11,9545	0,0045
7	$51e8_{-0,106}^{-0,060}$	50,917	0,0077
8	$51h8_{-0,046}$	50,977	0,0077
9	$31d9_{-0,142}^{-0,080}$	30,889	0,0103
10	$51d9_{-0,174}^{-0,1}$	50,863	0,0123
11	$11f8_{-0,043}^{-0,016}$	10,9705	0,0045
12	$32f8_{-0,064}^{-0,025}$	31,9555	0,0065
13	$11e9_{-0,075}^{-0,032}$	10,9465	0,0072
14	$32e9_{-0,112}^{-0,050}$	31,919	0,0103
15	$54f8_{-0,076}^{-0,030}$	53,947	0,0077

Приложение 2

ЗАДАНИЕ НА ФОРМИРОВАНИЕ ВЫБОРКИ - 25 ЦИКЛОВ ИЗМЕРЕНИЙ ПОСАДОЧНЫХ ДИАМЕТРОВ ПАРТИЙ ДЕТАЛЕЙ КЛАССА ВАЛОВ

Вариант	μ	σ
1	21	5
2	18	2
3	16	3
4	32	2
5	11	1
6	9	1
7	17	3
8	22	2
9	24	4
10	26	2
11	30	1
12	31	4
13	15	1
14	12	1
15	19	2

Приложение 3

ЗАДАНИЕ НА ФОРМИРОВАНИЕ ВЫБОРКИ ДЛЯ ПОСТРОЕНИЯ КОНТРОЛЬНОЙ КАРТЫ АЛЬТЕРНАТИВНЫХ ДАННЫХ

№ варианта	Значение p	Число испытаний
1	0,050	800
2	0,045	800
3	0,040	800
4	0,035	800
5	0,030	800
6	0,025	800
7	0,020	800
8	0,050	900
9	0,045	900
10	0,040	900
11	0,035	900
12	0,030	900
13	0,025	900
14	0,020	900
15	0,050	1000

Приложение 4

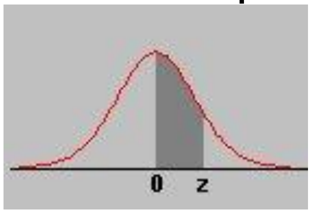


Таблица значений функции Гаусса $\boldsymbol{\varphi(x)} = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

X	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3031	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2089	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1316
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1067	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	055I
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0176	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0005	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0006	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0001
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Приложение 5

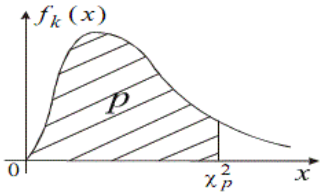


Таблица значений функции "хи–квадрат"

Число степеней свободы,	**Вероятность,**							
	.950	**.900**	**.750**	**.500**	**.250**	**.050**	**.010**	**.005**
1	0.00393	0.01579	0.10153	0.45494	1.32330	3.84146	6.63490	7.87944
2	0.10259	0.21072	0.57536	1.38629	2.77259	5.99146	9.21034	10.59663
3	0.35185	0.58437	1.21253	2.36597	4.10834	7.81473	11.34487	12.83816
4	0.71072	1.06362	1.92256	3.35669	5.38527	9.48773	13.27670	14.86026
5	1.14548	1.61031	2.67460	4.35146	6.62568	11.07050	15.08627	16.74960
6	1.63538	2.20413	3.45460	5.34812	7.84080	12.59159	16.81189	18.54758
7	2.16735	2.83311	4.25485	6.34581	9.03715	14.06714	18.47531	20.27774
8	2.73264	3.48954	5.07064	7.34412	10.21885	15.50731	20.09024	21.95495
9	3.32511	4.16816	5.89883	8.34283	11.38875	16.91898	21.66599	23.58935
10	3.94030	4.86518	6.73720	9.34182	12.54886	18.30704	23.20925	25.18818
11	4.57481	5.57778	7.58414	10.34100	13.70069	19.67514	24.72497	26.75685
12	5.22603	6.30380	8.43842	11.34032	14.84540	21.02607	26.21697	28.29952
13	5.89186	7.04150	9.29907	12.33976	15.98391	22.36203	27.68825	29.81947
14	6.57063	7.78953	10.16531	13.33927	17.11693	23.68479	29.14124	31.31935
15	7.26094	8.54676	11.03654	14.33886	18.24509	24.99579	30.57791	32.80132
16	7.96165	9.31224	11.91222	15.33850	19.36886	26.29623	31.99993	34.26719
17	8.67176	10.08519	12.79193	16.33818	20.48868	27.58711	33.40866	35.71847
18	9.39046	10.86494	13.67529	17.33790	21.60489	28.86930	34.80531	37.15645
19	10.11701	11.65091	14.56200	18.33765	22.71781	30.14353	36.19087	38.58226
20	10.85081	12.44261	15.45177	19.33743	23.82769	31.41043	37.56623	39.99685
21	11.59131	13.23960	16.34438	20.33723	24.93478	32.67057	38.93217	41.40106
22	12.33801	14.04149	17.23962	21.33704	26.03927	33.92444	40.28936	42.79565
23	13.09051	14.84796	18.13730	22.33688	27.14134	35.17246	41.63840	44.18128
24	13.84843	15.65868	19.03725	23.33673	28.24115	36.41503	42.97982	45.55851
25	14.61141	16.47341	19.93934	24.33659	29.33885	37.65248	44.31410	46.92789
26	15.37916	17.29188	20.84343	25.33646	30.43457	38.88514	45.64168	48.28988
27	16.15140	18.11390	21.74940	26.33634	31.52841	40.11327	46.96294	49.64492
28	16.92788	18.93924	22.65716	27.33623	32.62049	41.33714	48.27824	50.99338
29	17.70837	19.76774	23.56659	28.33613	33.71091	42.55697	49.58788	52.33562
30	18.49266	20.59923	24.47761	29.33603	34.79974	43.77297	50.89218	53.67196

Приложение 6

Коэффициенты для вычисления линий контрольных карт

Число наблюдений в подгруппе n	Коэффициенты для вычисления контрольных границ											Коэффициенты для вычисления центральной линии			
	A_1	A_2	A_3	B_3	B_4	B_5	B_6	D_1	D_2	D_3	D_4	C_4	$1/C_4$	d_2	$1/d_2$
2	2,121	1,880	2,659	0,000	3,267	0,000	2,606	0,000	3,686	0,000	3,267	0,7979	1,2533	1,128	0,8865
3	1,732	1,023	1,954	0,000	2,568	0,000	2,276	0,000	4,358	0,000	2,574	0,8886	1,1284	1,693	0,5907
4	1,500	0,729	1,628	0,000	2,266	0,000	2,088	0,000	4,696	0,000	2,282	0,9213	1,0854	2,059	0,4857
5	1,342	0,577	1,427	0,000	2,089	0,000	1,964	0,000	4,918	0,000	2,114	0,9400	1,0638	2,326	0,4299
6	1,225	0,483	1,287	0,030	1,970	0,029	1,874	0,000	5,078	0,000	2,004	0,9515	1,0510	2,534	0,3946
7	1,134	0,419	1,182	0,118	1,882	0,113	1,806	0,204	5,204	0,076	1,924	0,9594	1,0423	2,704	0,3698
8	1,061	0,373	1,099	0,185	1,815	0,179	1,751	0,388	5,306	0,136	1,864	0,9650	1,0363	2,847	0,3512
9	1,000	0,337	1,032	0,239	1,761	0,232	1,707	0,547	5,393	0,184	1,816	0,9693	1,0317	2,970	0,3367
10	0,949	0,308	0,975	0,284	1,716	0,276	1,669	0,687	5,469	0,223	1,777	0,9727	1,0281	3,078	0,3249
11	0,905	0,285	0,927	0,321	1,679	0,313	1,637	0,811	5,535	0,256	1,744	0,9754	1,0252	3,173	0,3152
12	0,866	0,266	0,886	0,354	1,646	0,346	1,610	0,922	5,594	0,283	1,717	0,9776	1,0229	3,258	0,3069
13	0,832	0,249	0,850	0,382	1,618	0,374	1,585	1,025	5,647	0,307	1,693	0,9794	1,0210	3,336	0,2998
14	0,802	0,235	0,817	0,406	1,594	0,399	1,563	1,118	5,696	0,328	1,672	0,9810	1,0194	3,407	0,2935
15	0,775	0,223	0,789	0,428	1,572	0,421	1,544	1,203	5,741	0,347	1,653	0,9823	1,0180	3,472	0,2880
16	0,750	0,212	0,763	0,448	1,552	0,440	1,526	1,282	5,782	0,363	1,637	0,9835	1,0168	3,532	0,2831
17	0,728	0,203	0,739	0,466	1,534	0,458	1,511	1,356	5,820	0,378	1,622	0,9845	1,0157	3,588	0,2784
18	0,707	0,194	0,718	0,482	1,518	0,475	1,496	1,424	5,856	0,391	1,608	0,9854	1,0148	3,640	0,2747
19	0,688	0,187	0,698	0,497	1,503	0,490	1,483	1,487	5,891	0,403	1,597	0,9862	1,0140	3,689	0,2711
20	0,671	0,180	0,680	0,510	1,490	0,504	1,470	1,549	5,921	0,415	1,585	0,9869	1,0133	3,735	0,2677
21	0,655	0,173	0,663	0,523	1,477	0,516	1,459	1,605	5,951	0,425	1,575	0,9876	1,0126	3,778	0,2647
22	0,640	0,167	0,647	0,534	1,466	0,528	1,448	1,659	5,979	0,434	1,566	0,9882	1,0119	3,819	0,2618
23	0,626	0,162	0,633	0,545	1,455	0,539	1,438	1,710	6,006	0,443	1,557	0,9887	1,0114	3,858	0,2592
24	0,612	0,157	0,619	0,555	1,445	0,549	1,429	1,759	6,031	0,451	1,548	0,9892	1,0109	3,895	0,2567
25	0,600	0,153	0,606	0,565	1,434	0,559	1,420	1,806	6,056	0,459	1,541	0,9896	1,0105	3,931	0,2544

Примечание - Источник ASTM, Philadelphia, PA, USA.

Приложение 7

Таблица значений F-критерия Фишера при уровне значимости

k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
1	161,5	199,5	215,7	224,6	230,2	233,9	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1

Приложение 8

Таблица значений функции Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{x^2}{2}}dx$

x	Ф*(x)*	*x*	Ф*(x)*	*x*	Ф*(x)*	*x*	Ф*(x)*
0,00	0,0000	0,44	0,1700	0,88	0,3106	1,32	0,4066
0,01	0,0040	0,45	0,1736	0,89	0,3133	1,33	0,4082
0,02	0,0080	0,46	0,1772	0,90	0,3159	1,34	0,4099
0,03	0,0120	0,47	0,1808	0,91	0,3186	1,35	0,4116
0,04	0,0160	0,48	0,1844	0,92	0,3212	1,36	0,4131
0,05	0,0199	0,49	0,1879	0,93	0,3238	1,37	0,4147
0,06	0,0239	0,59	0,1915	0,94	0,3264	1,38	0,4162
0,07	0,0279	0,51	0,1950	0,95	0,3289	1,39	0,4177
0,08	0,0319	0,52	0,1985	0,96	0,3315	1,40	0,4192
0,09	0,0359	0,53	0,2019	0,97	0,3340	1,41	0,4207
0,10	0,0398	0,54	0,2054	0,98	0,3365	1,42	0,4222
0,11	0,0438	0,55	0,2088	0,99	0,3369	1,43	0,4236
0,12	0,0478	0,56	0,2123	1,00	0,3413	1,44	0,4261
0,13	0,0517	0,57	0,2157	1,01	0,3438	1,45	0,4265
0,14	0,0557	0,58	0,2190	1,02	0,3461	1,46	0,4279
0,15	0,0596	0,59	0,2224	1,03	0,3485	1,47	0,4292
0,16	0,0636	0,60	0,2257	1,04	0,3508	1,48	0,4306
0,17	0,0675	0,61	0,2291	1,05	0,3531	1,49	0,4319
0,18	0,0724	0,62	0,2224	1,06	0,3554	1,50	0,4332
0,19	0,0753	0,63	0,2357	1,07	0,3677	1,51	0,4345
0,20	0,0793	0,64	0,2389	1,08	0,3599	1,52	0,4357
0,21	0,0832	0,65	0,2422	1,09	0,3621	1,53	0,4370
0,22	0,0871	0,66	0,2454	1,10	0,3043	1,54	0 4382
0,23	0,0910	0,67	0,2486	1,11	0,3665	1,55	0,4394
0,24	0,0948	0,68	0,2517	1,12	0,3686	1,56	0,4406
0,25	0,0987	0,69	0,2549	1,13	0,3708	1,57	0,4418
0,26	0,1026	0,70	0,2580	1,14	0,3729	1,58	0,4429
0,27	0,1064	0,71	0,2611	1,15	0,3749	1,59	0,4441
0,28	0,1103	0,72	0,2642	1,l6	0,3770	1,60	0,4452
0,29	0,1141	0,73	0,2673	1,17	0,3790	1,61	0,4463
0,30	0,1179	0,74	0,2703	1,18	0,3810	1,62	0,4474
0,31	0,1217	0,75	0,2734	1,19	0,2830	163	0,4484
0,38	0,1265	0,76	0,2764	1,20	0,3849	1,64	0,4496
0,33	0,1293	0,77	0,2794	1,21	0,3869	1,65	0,4505
0,34	0,1331	0,78	0,2823	1,22	0,3883	1,66	0,4615
0,35	0,1368	0,79	0,2852	1,23	0,3907	1,67	0,4625
0,36	0,1406	0,80	0,2881	1,24	0,3925	1,68	0,4636
0,37	0,1443	0,81	0,2910	1,25	0,3944	1 69	0,4645
0,38	0,1480	0,82	02839	1,26	0,3961	1,70	0,4654
0,39	0,1517	0,83	0,2967	1,27	0,3980	1,71	0,4664
0,40	0,1664	0,84	0,8996	1,28	0,8997	1,72	0,4673
0,41	0,1691	0,85	0,3023	1,29	0,4015	1,73	0,4682
0,42	0,1688	0,86	0,3061	1,30	0,4032	1,74	0,4691
0,43	0,1664	0,87	0,3078	1,31	0,4049	1,75	0,4699

1,76	0,4608	1,97	0,4756	2,36	0,4909	2,76	0,4971
1,77	0,4616	1,98	0,4761	2,38	0,4913	2,78	0,4973
1,78	0,4625	1,99	0,4767	2,40	0,4918	2,80	0,4974
1,79	0,4633	2,00	0,4772	2,42	0,4922	2,82	0,4976
1,80	0,4641	2,02	0,4783	2,44	0,4927	2,84	0,4977
1,81	0,4649	2,04	0,4793	2,46	0,4931	2,86	0,4979
1,82	0,4656	2,06	0,4803	2,48	0,4934	2,88	0,4980
1,83	0,4664	2,08	0,4812	2,50	0,4938	2,90	0,4981
1,84	0,4671	2,10	0,4821	2,52	0,4941	2,92	0,4982
1,86	0,4678	2,12	0,4830	2,54	0,4945	2,94	0,4984
1,86	0,4686	2,14	0,4838	2,56	0,4948	2,96	0,4985
1,87	0,4693	2,16	0,4846	2,58	0,4951	2,98	0,4986
1,88	0,4699	2,18	0,4854	2,60	0,4953	3,00	0,49865
1,89	0,4706	2,20	0,4861	2,62	0,4956	3,20	0,49931
1,90	0,4713	2,22	0,4868	2,64	0,4959	3,40	0,49966
1,91	0,4719	2,24	0,4875	2,66	0,4961	3,60	0,499841
1,92	0,4726	2,26	0,4881	2,68	0,4963	3,80	0,499928
1,93	0,4732	2,28	0,4887	2,70	0,4968	4,00	0,499968
1,94	0,4738	2,30	0,4893	2,72	0,4967	4,50	0,499997
1,95	0,4744	2,32	0,4898	2,74	0,4969	5,00	0,499997
1,98	0,4761	2,34	0,4904				

Приложение 9

Значение коэффициентов Стьюдентаtпри вероятности

$$P\{|t| < t_p\} = 2\int_0^{t_p} p(t,n)dt$$

Число степеней свободы n-1	Вероятность P			
	0,90	0,95	0,98	0,99
1	6,314	12,706	31,821	63,657
2	2,920	4,303	6,965	9925
3	2,353	3,182	4,541	5,841
4	2,132	2,776	3,747	4,604
5	2,015	2,571	3,365	4,032
6	1,943	2,447	3,143	3,707
7	1,895	2,365	2,998	3,499
8	1,860	2,306	2,896	3,355
9	1,833	2,262	2,821	3,250
10	1,812	2,228	2,764	3,169
11	1,796	2,201	2,718	3,106
12	1,782	2,179	2,681	3,055
13	1,771	2,160	2,650	3,012
14	1,761	2,145	2,624	2,977
15	1,753	2,131	2,602	2,947
16	1,746	2,120	2,583	2,921
17	1,740	2,110	2,567	2,898
18	1,734	2,101	2,552	2,878
19	1,729	2,093	2,539	2,861
20	1,725	2,086	2,528	2,845
21	1,721	2,080	2,518	2,831
22	1,717	2,074	2,508	2,819
23	1,714	2,069	2,500	2,807
24	1,711	2,064	2,492	2,707
25	1,708	2,060	2,485	2,787
26	1,706	2,056	2,479	2,779
27	1,703	2,052	2,473	2,771
28	1,701	2,048	2,467	2,763
29	1,699	2,045	2,462	2,756
30	1,697	2,042	2,457	2,750
∞	1,64485	1,95996	2,32634	2,57582

Приложение 10

Критические значения критерия U Вилкоксона-Манна-Уитни
для уровней статистической значимости P 0,05

n_1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
n_2	**P = 0,05**																		
3	-	0																	
4	-	0	1																
5	0	1	2	4															
6	0	2	3	5	7														
7	0	2	4	6	8	11													
8	1	3	5	8	10	13	15												
9	1	4	6	9	12	15	18	21											
10	1	4	7	11	14	17	20	24	27										
11	1	5	8	12	16	19	23	27	31	34									
12	2	5	9	13	17	21	26	30	34	38	42								
13	2	6	10	15	19	24	28	33	37	42	47	51							
14	3	7	11	16	21	26	31	36	41	46	51	56	61						
15	3	7	12	18	23	28	33	39	44	50	55	61	66	72					
16	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83				
17	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96			
18	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109		
19	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	
20	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138