

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Локтионова Оксана Геннадьевна

Должность: проректор по учебной работе

Дата подписания: 29.05.2023 12:24:58

Уникальный программный ключ:

0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

МИНОБРАЗОВАНИЯ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования

«Юго-Западный государственный университет»

(ЮЗГУ)

Кафедра дизайна и индустрии моды

УТВЕРЖДАЮ

Проректор по учебной работе



« 26 » 5



**ПРИМЕНЕНИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ
КОНТРОЛЯ И УПРАВЛЕНИЯ КАЧЕСТВОМ НА
ПРОИЗВОДСТВЕ**

Методические указания по выполнению лабораторных работ и
самостоятельной работы

Курск 2023

УДК 650.012

Составители: С.В. Ходыревская

Рецензент

Доктор технических наук, доцент *В.В. Куц*

Применение статистических методов контроля и управления качеством на производстве: методические указания по выполнению лабораторных работ и самостоятельной работы / Минобрнауки России, Юго-Зап. гос. ун-т; сост.: С.В. Ходыревская. – Курск, 2023. – 142 с.:– Библиогр.: с. 130.

Содержат сведения о статистических методах контроля и управления качеством. Рассмотрены примеры проведения статистического анализа в программе *STATISTICA* и среде *LibreOffice*. Приведены задания для самостоятельного выполнения, контрольные вопросы, а также тест для самоконтроля.

Методические указания соответствуют Федеральному Государственному образовательному стандарту направления подготовки 27.04.02 Управление качеством.

Предназначены для магистров всех форм обучения направлений подготовки, изучающих дисциплину «Применение статистических методов контроля и управления качеством на производстве».

Текст печатается в авторской редакции

Подписано в печать . Формат 60x84 1/16.

Усл.печ. л. 8,25. Уч.-изд. л. 7,47.

Тираж 100 экз. Заказ *456* Бесплатно.

Юго-Западный государственный университет.

305040, г. Курск, ул. 50 лет Октября, 94.

Содержание

Лабораторная работа №1 Сбор данных с помощью контрольных листков	4
Контрольные вопросы	8
Лабораторная работа №2 Исследование распределения статистических данных для анализа качества и пригодности процесса	9
Контрольные вопросы	38
Лабораторная работа №3 Выявление основных причин появления проблемы на основании анализа диаграммы Парето ...	39
Контрольные вопросы	47
Лабораторная работа №4 Применение корреляционного и регрессионного анализа для установления причин дефектности продукции	48
Контрольные вопросы	81
Лабораторная работа №5 Применение контрольных карт Шухарта для анализа стабильности и управления технологическим процессом	83
Контрольные вопросы	102
Лабораторная работа №6 Расчет измерителей качества.....	104
Контрольные вопросы	109
Лабораторная работа №7 Методы анализа влияния факторов...110	110
Контрольные вопросы	125
Тестовые задания для самоконтроля	127
Библиографический список	130
Приложение А – Контрольный листок для регистрации распределения измеряемого параметра в ходе производственного процесса	132
Приложение Б – Контрольный листок видов дефектов	133
Приложение В – Контрольный листок локализации дефектов	134
Приложение Г –Контрольный листок причин дефектов	135
Приложение Д –Таблица критических точек распределения χ^2 (хи-квадрат) критерия Пирсона	136
Приложение Е – Таблица для определения уровня несоответствий продукции	137
Приложение Ж – Измерители качества и целевые показатели	138
Приложение И – Таблица преобразований для 6 сигма.....	142

Лабораторная работа №1

Сбор данных с помощью контрольных листков

Цель работы: получить сведения о порядке сбора данных, назначении и видах контрольных листков и освоить методику выбора контрольного листка и приобрести практические навыки его построения для сбора статистических данных.

Задания для самостоятельного выполнения:

Задание 1. Деталь обрабатывают на станке и в результате замеров получены размеры внутреннего диаметра вала, выданные преподавателем. Нормативно-технической документацией предусмотрен допуск равный $0,750 \pm 0,01$ мм. Разработать контрольный листок регистрации распределения измеряемого параметра и контрольный листок регистрации данных.

Задание 2. Поверхность детали параллельно обрабатывается на 5 однотипных станках, каждый из которых характеризуется различной точностью и уровнем настройки. На основании данных полученных в течение смены разработать контрольный листок для учета дефектов.

Краткие теоретические сведения

Порядок сбора данных

Прежде чем начать собирать данные необходимо решить, для чего мы это делаем.

Цель сбора данных в процессе контроля качества состоит в следующем: контроль и регулирование процесса; анализ отклонений от установленных требований; контроль выхода процесса.

Когда цель сбора данных установлена, она становится основой для определения характера сравнений, которые надлежит произвести, и типа данных которые нужно собрать.

Важно в процессе сбора тщательно упорядочить данные, чтобы облегчить их последующую обработку.

Во-первых, надо четко зарегистрировать источник данных (без такой регистрации данные окажутся «мертвыми»). Очень часто, несмотря на то, что была потрачена целая неделя на сбор данных о показателях качества, из них можно извлечь мало полезной информации, поскольку не зафиксированы день недели,

когда собирались данные, оборудование, на котором производилась работа, рабочий, делавший операцию, партия используемых материалов и т.д.

Во-вторых, данные надо регистрировать таким образом, чтобы их было легко использовать. Поскольку данные часто применяются для вычисления статистических характеристик (средние значения и размах), то лучше их записывать так, чтобы облегчить эти вычисления.

Если данные требуется собирать постоянно, то надо заранее разработать стандартные формы регистрации данных. Для этого используются различные виды контрольных листков.

Контрольный листок – бумажный бланк, на котором заранее напечатаны контрольные параметры, с тем, чтобы можно было легко и точно записать данные измерений. Его главное назначение: облегчить процесс сбора данных; автоматически упорядочить данные для облегчения их дальнейшего использования.

Обычно несколько людей занимаются регистрацией и обработкой данных, поэтому возникает большая вероятность появления ошибок в процессе записи. Контрольный листок и помогает свести эти ошибки к минимуму. На него заносят данные с помощью пометок или простых символов, что позволяет автоматически упорядочить данные без их последующего переписывания от руки.

Виды контрольных листков

Контрольный листок для регистрации распределения измеряемого параметра в ходе производственного процесса

Предположим, что мы хотим выявить изменения в размерах некоторой детали, подвергающейся механической обработке, причем размер, указанный в чертеже $8,300 \pm 0,008$. Для получения распределения значений этого показателя в ходе процесса обычно используются гистограммы. На основе гистограммы вычисляются среднее значение и дисперсия, исследуется также и форма кривой распределения. Чтобы построить гистограмму, надо затратить немало труда на сбор большого числа данных и на представление частотного распределения в графической форме. Проще классифицировать данные в момент сбора.

В Приложении А показан бланк, который заранее готовится

для этой цели. Каждый раз, когда производится замер, в соответствующую клеточку ставится крест, так что к концу измерений гистограмма готова. Если нужно произвести расслоение с использованием одного контрольного листка, лучше для пометок брать карандаши разного цвета, чтобы разница проявлялась наглядно.

Контрольный листок для регистрации видов дефектов

Рассмотрим контрольный листок, используемый в процессе приемочного контроля одной штампованной пластиковой детали (см. Приложение Б). Всякий раз, когда контролер обнаруживает дефект, он делает в контрольном листке пометку. В конце рабочего дня он может быстро подсчитать число и разновидности встретившихся дефектов.

Просто знание числа дефектов не позволяет принять решение о корректирующих мерах, но если использовать подобный контрольный листок, то он даст необходимую информацию для совершенствования процесса, поскольку показывает какие виды дефектов встречаются часто, а какие – нет.

Используя подобный контрольный листок, нельзя потом произвести расслоение данных, разбив их, например, на утреннюю и вечернюю смены, как это было возможно в случае с предыдущим рассмотренным листком. Поэтому если расслоение необходимо, то при разработке бланка это нужно предусмотреть заранее.

Контрольный листок локализации дефектов

Во всех видах продукции обнаруживаются внешние дефекты, такие, как царапины или грязь, и на многих предприятиях предпринимаются разные меры для их сокращения. Большую роль в решении этой проблемы играют контрольные листки локализации дефектов. В листках такого рода есть эскизы или схемы, на которых делаются пометки, так что можно пронаблюдать расположение дефектов. В Приложении В показан такой листок, используемый на машиностроительном заводе для входного контроля отливок. Дефект – раковины в отливках, и раньше поставщика информировали только о приемке или браковке партий и числе дефектов в партии, но качество не ухудшалось. После того как были введены листки локализации дефектов, на которых фиксировалось возможное месторасположение раковин, качество

отливок резко улучшилось, поскольку облегчился поиск причин дефектов. Такого типа контрольные листки необходимы для диагноза процесса, поскольку причины дефектов часто можно найти, исследуя места их возникновения и наблюдая процесс в поисках объяснений, почему дефекты концентрируются именно в этих местах.

Контрольный листок причин дефектов

Листок локализации дефектов, описанный в предыдущем примере, используется для привлечения внимания к местам возникновения дефектов. Кроме того, контрольные листки применяются иногда для дальнейшей стратификации с целью определения причин дефектов. Вообще говоря, большинство исследований, нацеленных на поиск причин дефектов, предполагают сравнение данных о причинах с соответствующими данными о последствиях дефектов, сопоставляя их в строго установленном порядке для последующего анализа путем стратификации или с использованием диаграмм рассеивания. Но если случай достаточно простой, то можно фиксировать необходимые данные прямо в контрольном листке.

Например, в Приложении Г представлен контрольный листок для регистрации дефектов с учетом станков, рабочих, дней изготовления и типов дефектов. С первого взгляда видно, что рабочий В допускает много дефектов. Все рабочие допустили много дефектов в среду. Исследование причин дефектов показало, что рабочий В недостаточно часто менял штампы, а причиной дефектов, допущенных в среду, послужило плохое качество исходного сырья.

С той же целью можно использовать контрольные листки в виде диаграмм причин и результатов, которые понятны каждому рабочему. Такие диаграммы вычерчиваются заранее, и когда становится ясна причина или условия допущения дефекта, то у соответствующей стрелочки делается пометка. По этой диаграмме можно определить, устранению каких причин важно отдавать предпочтение

Помимо описанных примеров, существует большое число других контрольных листков и форм, используемых на производстве. Они разрабатываются, прежде всего, с учетом цели

сбора данных, а потом создаются модификации, подходящие для поставленных целей и облегчающие сбор и регистрацию данных.

Порядок выполнения работы

Получив у преподавателя исходные данные для выполнения лабораторной работы, студент изучает порядок сбора данных и различные виды контрольных листков.

Затем приступает непосредственно к выполнению задания на компьютере.

Содержание отчета

Отчет по лабораторной работе должен содержать следующие пункты:

- название лабораторной работы;
- цель работы;
- краткие теоретические сведения;
- краткое описание хода выполнения работы;
- индивидуальное задание для выполнения лабораторной работы;
- результаты выполнения работы: контрольный листок регистрации распределения измеряемого параметра; контрольный листок регистрации данных и контрольный листок учета дефектов;
- выводы

Контрольные вопросы

1. Что такое контрольный листок?
2. Каков порядок сбора информации?
3. Каков порядок сбора информации с помощью контрольных листков?
4. Каковы основные преимущества сбора данных с помощью контрольных листков?
5. Какие типы контрольных листов Вам известны?
6. Где и как применяются контрольные листки?

Лабораторная работа №2

Исследование распределения статистических данных для анализа качества и пригодности процесса

Цель работы: исследовать распределения статистических данных с помощью проверки нормальности распределения, построения гистограмм в программе *STATISTICA* или *LibreOffice*, а также провести анализ точности и пригодности процесса.

Задание для самостоятельного выполнения:

Для исследования распределения диаметров стальных осей, полученных на токарном станке, были измерены диаметры 100 осей, полученные данные выдаются преподавателем. Провести проверку нормальности распределения, построить гистограмму и провести анализ точности и пригодности процесса обработки стальных осей на токарном станке.

Краткие теоретические сведения

Проверка нормальности распределения с помощью построения статистической функции распределения

При малом числе наблюдений для оценки нормальности можно воспользоваться статистической функцией распределения результатов наблюдений. Для ее построения, полученные в процессе эксперимента результаты, группируют в вариационный ряд $X_{(1)}^*$; $X_{(2)}^*$; ...; $X_{(n)}^*$, члены которого располагаются в порядке их возрастания, так что всегда $X_{(1)}^* \leq X_{(2)}^* \leq \dots \leq X_{(n)}^*$. Статистическую функцию распределения $F_n(X_k)$ определяют по формуле

$$F_n(X_k) = \frac{k}{n+1}; k = 1, 2, \dots, n. \quad (2.1)$$

$F_n(X)$ представляет собой ступенчатую линию, скачки которой соответствуют значениям членов вариационного ряда. Каждый скачок равен $\frac{1}{n+1}$, если все n членов ряда различны. Если же для некоторого k имеет место равенство $X_{(k)} = X_{(k+1)} = \dots = X_{(k+i)}$, то $F_n(X)$ в точке $X = X_{(k)}$ возрастает на $\frac{i}{n+1}$, где i — число равных между собой членов ряда.

Если число наблюдений безгранично увеличивать, то

статистическая функция распределения сходится по вероятности к истинной функции $F_x(X)$.

Для проверки нормальности распределения результатов наблюдений по таблицам находят значения Z_k , соответствующие полученным значениям $F_n(X_k)$ статистической функции распределения: $\Phi(Z_k) = F_n(X_k)$. Переменная Z определяется через результаты наблюдений как $Z_k = \frac{X_k - m_x}{\sigma_x}$, и если в координатах Z ; X

нанести точки Z_k ; X_k , то при нормальном распределении они должны расположиться вдоль одной прямой линии (см. рисунок 2.1,а). Если же в результате такого построения получится некоторая кривая линия (см. рисунок 2.1,б), то гипотезу о нормальности распределения придется отвергнуть как противоречащую опытным данным.

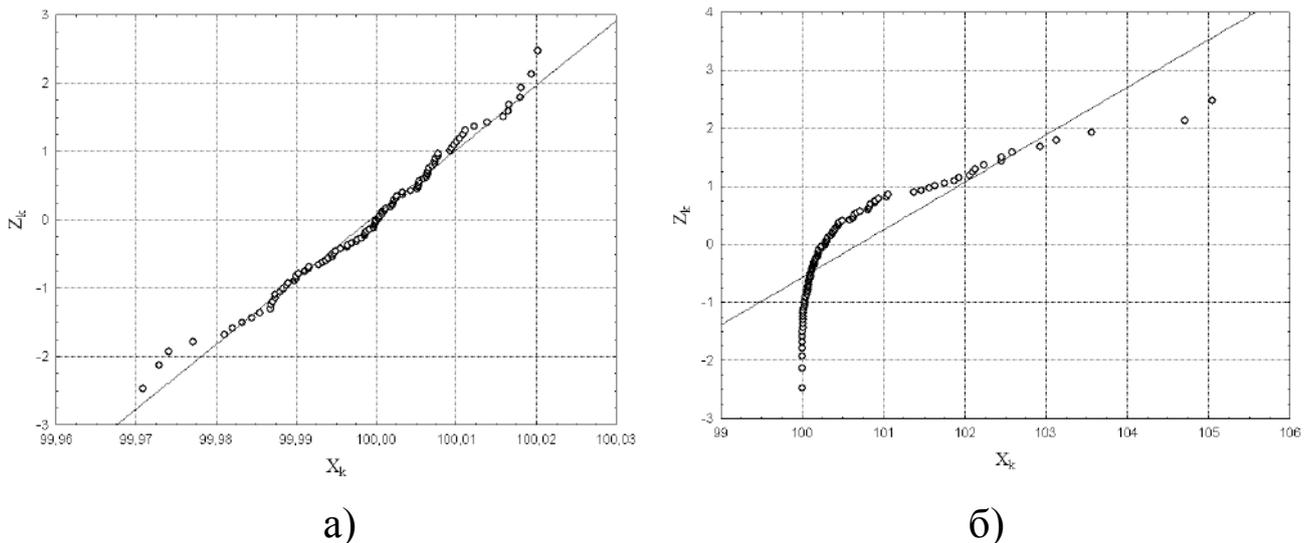


Рисунок 2.1 – Статистические функции распределения

Проверка нормальности распределения согласно критерию χ^2

Проверка нормальности распределения согласно критерию χ^2 сводится к следующему.

1. Данные наблюдений группируют по интервалам.

Весь диапазон полученных результатов наблюдений $X_{\max} - X_{\min}$ разделяют на r интервалов шириной $\Delta X_j = (X_{\max} - X_{\min})/r$, ($j=1,2,\dots, r$) и подсчитывают частоты, m_j , равные числу результатов, лежащих в каждом j -м интервале, т. е. меньших или равных его правой и больших левой границы. Число r интервалов выбирается

нечетным в зависимости от числа наблюдений согласно следующим рекомендациям таблицы 2.1:

Таблица 2.1

n	r
40—100	7—9
100—500	9—13
500—1000	11—17
1000—10000	13—23

Отношения

$$P_j^* = \frac{m_j}{n} \quad (2.2)$$

где n — общее число наблюдений, называются частотами и представляют собой статистические оценки вероятностей попадания результата наблюдений в j -й интервал. Распределение частостей по интервалам образует статистическое распределение результатов наблюдений.

Если разделить частость на длину интервала, то получим величины

$$p_j^* = \frac{1}{\Delta X_j} P_j^* = \frac{m_j}{n \Delta X_j} \quad (2.3)$$

являющиеся оценками средней плотности распределения в интервале ΔX_j .

Если в некоторые интервалы попадает меньше пяти наблюдений, то такие интервалы объединяют с соседними. При этом число степеней свободы k уменьшается на единицу.

2. Вычисляют среднее арифметическое \bar{X}

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.4)$$

и точечную оценку среднего квадратического отклонения результата наблюдений s_x

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}, \quad (2.5)$$

которые принимают в качестве параметров теоретического нормального распределения с плотностью $p_x(x)$.

3. Для каждого интервала находят вероятности попадания в

них результатов наблюдений приближенно как произведение плотности теоретического распределения в середине интервала на его длину:

$$P_j \approx p_X \left(\frac{X_j + X_{j+1}}{2} \right) \Delta X_j. \quad (2.6)$$

4. Для каждого интервала вычисляют величину меры расхождения χ_j^2 ($j=1,2,\dots,r$)

$$\chi_j^2 = \frac{(m_j - nP_j)^2}{nP_j} \quad (2.7)$$

и суммируют их по всем j , в результате чего получают меру расхождения χ_k^2

$$\chi_k^2 = \sum_{j=1}^r \frac{n}{P_j} (P_j^* - P_j)^2 = \sum_{j=1}^r \frac{(m_j - nP_j)^2}{nP_j} = \sum_{i=1}^r \chi_j^2. \quad (2.8)$$

5. Определяют число степеней свободы $k = r - 3$ и, задаваясь уровнем значимости $q = 1 - \alpha$, находят по справочным таблицам значения $\chi_{k;q}^2$ и $\chi_{k;1-q}^2$. Если $\chi_{k;q}^2 < \chi_k^2 \leq \chi_{k;1-q}^2$, то распределение результатов наблюдений считают нормальным.

Отметим, что под уровнем значимости q понимается вероятность того, что результаты измерений, признанные распределенными не по нормальному закону в действительности распределены нормально (т.е. ошибка первого рода).

Визуальная оценка распределения

Полезную информацию о возможном характере распределения можно получить, взглянув на рисунки 2.2-2.8. Формы, представленные на этих рисунках, типичны, и ими можно воспользоваться как образцами при анализе процессов.

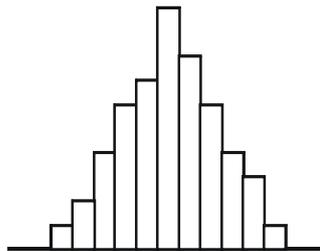


Рисунок 2.2 – Обычный тип гистограммы
Обычный тип (симметричный). Гистограмма с таким

распределением встречается чаще всего. Она указывает на стабильность процесса.

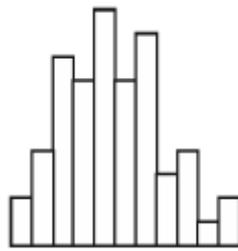


Рисунок 2.3 – Мультимодальный тип гистограммы

Гребенка (мультимодальный тип). Здесь классы через один имеют более низкие частоты. Такая форма встречается, когда число единичных наблюдений, попадающих в класс, колеблется от класса к классу или, когда действует определенное правило округления данных.

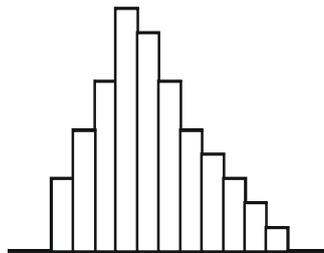


Рисунок 2.4 – Гистограмма положительно скошенного распределения

Положительно (отрицательно) скошенное распределение. Среднее значение гистограммы локализуется слева (справа) от центра размаха. Частоты довольно резко спадают при движении влево (вправо) и, наоборот, медленно вправо (влево). Такая (асимметричная) форма встречается, когда невозможно получить значения ниже определенного, например для диаметра деталей и т.д.

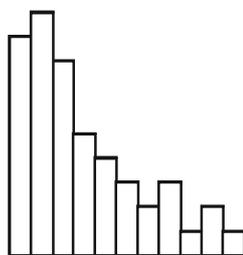


Рисунок 2.5 – Гистограмма распределения с обрывом слева

Распределение с обрывом слева (справа). Это одна из тех форм, которые часто встречаются при 100%-ном контроле изделий из-за плохой воспроизводимости процесса, а также когда, например, отобраны и исключены из партии все изделия с

параметрами ниже контрольного нормативы (или выше, или и те и другие).

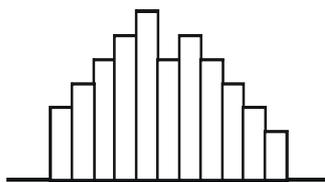


Рисунок 2.6 – Гистограмма равномерного распределения Плато (равномерное и прямоугольное распределение). Такая гистограмма получается в случаях, когда объединяются несколько распределений, в которых средние значения имеют небольшую разницу между собой. Анализ такой гистограммы целесообразно проводить, используя метод расслоения.

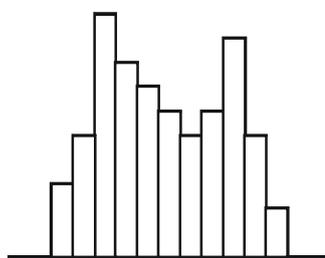


Рисунок 2.7 – Двухпиковый тип гистограммы
Двухпиковый тип (бимодальный тип). Такая форма встречается, когда смешиваются два распределения с далеко отстоящими средними значениями, например, в случае наличия разницы между двумя видами материалов, двумя операторами и т.д. В этом случае можно провести расслоение по двум видам фактора, исследовать причины различия и принять соответствующие меры для его устранения.

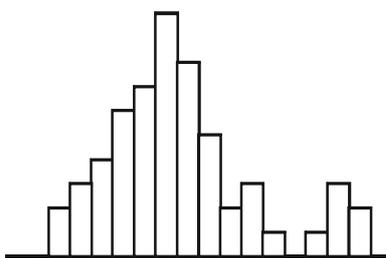


Рисунок 2.8 – Гистограмма распределения с изолированным пиком

Распределение с изолированным пиком. Рядом с распределением обычного типа появляется маленький изолированный пик. Это форма появляется при наличии малых включений данных из другого распределения, появления ошибки измерения или просто включения данных из другого процесса. По

результатам анализа гистограммы дают заключение о необходимости настройки измерительного прибора или срочного осуществления контроля процесса.

Визуальная оценка расположения статистических данных в пределах поля допуска

Чтобы оценить адекватность процесса требованиям потребителя, мы должны сравнить качество процесса с полем допуска. Если имеется допуск, то на гистограмму наносят верхнюю (U) и нижнюю (L) его границы в виде линий, перпендикулярных оси абсцисс, чтобы сравнить распределение параметров качества процесса с этими границами. В технических условиях (ТУ) может быть указано номинальное значение измеряемого параметра качества. Номинальное значение представляет собой математическое ожидание, т.е. среднее значение измеряемого показателя качества, которое можно рассматривать как генеральную характеристику, а совокупность всех полученных показателей качества, как генеральную совокупность. В ТУ также установлены допустимые пределы изменения данного показателя. При этом ширина поля допуска T определяется как величина, равная разности значений верхней и нижней границ допуска, т.е. $T = U - L$.

Существует пять типичных случаев расположения статистических данных в пределах поля допуска, показанных на рисунках 2.9-2.10.

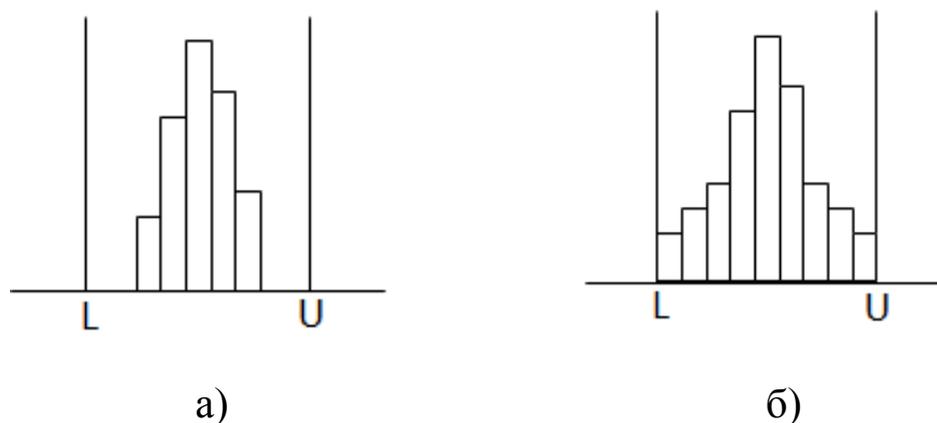


Рисунок 2.9 – Гистограмма удовлетворяет допуску

Если гистограмма удовлетворяет допуску (см. рисунок 2.9), то в случаях:

- а) поддержание существующего состояния – это все, что

требуется, поскольку гистограмма вполне соответствует допускам;

б) допуски удовлетворяются, но нет никакого запаса, поэтому необходимо сократить разброс до меньшего значения.

Когда гистограмма не удовлетворяет допуску (см. рисунок 2.10), то в случаях:

а) необходимо добиться смещения среднего ближе к центру поля допуска;

б) требуются действия, направленные на снижение вариации;

в) одновременно требуются меры, описанные в пунктах а) и б).

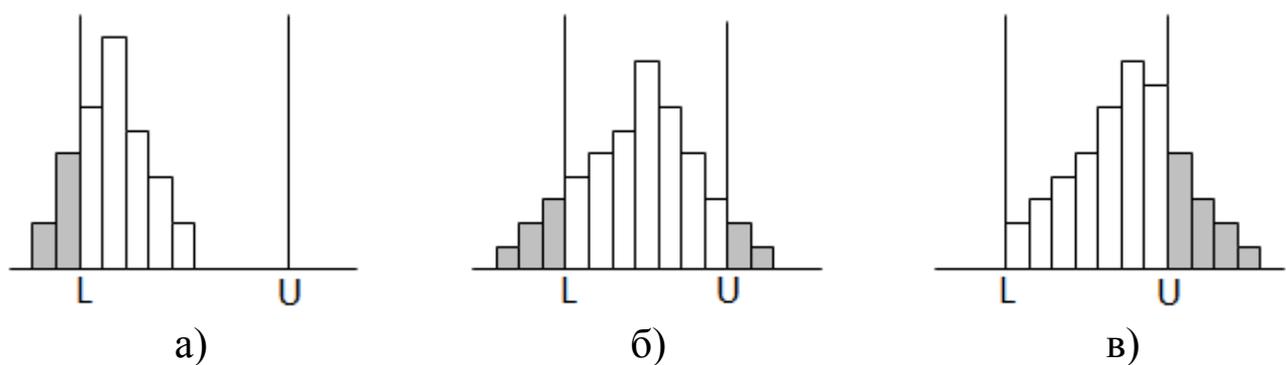


Рисунок 2.10 – Гистограмма не удовлетворяет допуску

Анализ пригодности процесса

Когда выясняется, что гистограмма следует нормальному распределению, часто предпринимается исследование пригодности процесса, т.е. определяется неизменность основных параметров процесса: среднего значения \bar{x} и стандартного отклонения во времени σ . Оно важно при оценке того, сможет ли процесс пересечь границы поля допуска или нет и появления в связи с этим несоответствия требованиям потребителя. Если допустить, что процесс имеет нормальное распределение, то можно сразу же определить процент дефектов, оказавшихся за данными границами допуска при данных параметрах (\bar{x} , s). Но более полезно оценить процесс с помощью оценки параметров пригодности процесса в соответствии с ГОСТ Р ИСО 22514-1-2015 Статистические методы. Управление процессами. Часть 1. Общие принципы.

Показатель пригодности процесса – статистический показатель выходной характеристики процесса, используемый для описания процесса, пребывание которого в состоянии статистической управляемости не подтверждено.

Этап 1. Расчет индексов пригодности процесса

Анализ пригодности процесса следует проводить с помощью P_p – индекса пригодности. Индекс пригодности – величина, которая характеризующая меру пригодности процесса относительно установленного поля допуска. При нормальном распределении он определяется по формуле:

$$P_p = \frac{U - L}{6s_t} \quad (2.9)$$

Для нормального распределения оценка среднего квадратического отклонения по одной выборке размера n имеет вид:

$$s_t = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x}_t)^2} \quad (2.10)$$

где $\bar{x}_t = \frac{1}{n} \sum x_i$

Верхний индекс пригодности процесса P_{pkU} – индекс пригодности процесса относительно верхней границы поля допуска U , при нормальном распределении определяется по формуле:

$$P_{pkU} = \frac{U - \bar{X}_t}{3s_t} \quad (2.11)$$

Нижний индекс пригодности процесса P_{pkL} – индекс пригодности процесса относительно нижней границы поля допуска L , при нормальном распределении определяется по формуле:

$$P_{pkL} = \frac{\bar{X}_t - L}{3s_t} \quad (2.12)$$

Меньший индекс пригодности процесса определяется по формуле:

$$P_{pk} = \min \{ P_{pkU}; P_{pkL} \} \quad (2.13)$$

Этот индекс иногда называют «критическим».

Иногда в спецификации указана только одна граница, например максимальное (или минимальное) значение. В этом случае можно вычислить верхний индекс P_{pk} только на основе

верхнего (или нижнего) индекса пригодности процесса.

Возможна ситуация, когда границы поля допуска заданы, а целевое значение отличается от середины интервала, заданного в спецификации, и целевое значение является предпочтительным (или лучшим) значением. В этом случае можно вычислить только верхний и нижний, но не меньший индексы пригодности процесса. Для верхних и нижних индексов могут быть установлены различные требования.

Этап 2. Анализ пригодности процесса

Исследование пригодности процесса с помощью P_p позволяет оценить качество процесса в соответствии с требованиями потребителя. Чем больше величина P_p , тем выше качество процесса и тем меньше вероятность несоответствия выхода процесса ожиданиям потребителя.

Пригодность технологического процесса оценивают, исходя из следующих критериев:

1. В случае, когда $P_p \geq 1,67$, ширина интервала между контрольными нормативами не менее чем в 10 раз превышает стандартное отклонение σ ; разброс параметров изделия невелик, появление брака не угрожает. Целесообразно несколько понизить класс исходного сырья (материалов, комплектующих) и упростить контроль процесса, что приведет к снижению себестоимости продукции (см. рисунок 2.11а).

2. В случае, когда $1,67 > P_p \geq 1,33$, ширина интервала между контрольными нормативами в 8-10 раз превышает стандартное отклонение σ . Идеальное состояние процесса (см. рисунок 2.11б).

3. В случае, когда $1,33 > P_p \geq 1,00$, ширина интервала между контрольными нормативами в 6-8 раз превышает стандартное отклонение σ . Когда показатель P_p близок к 1, вероятность появления брака составляет 0,27%, поэтому необходимо усилить контроль процесса, провести анализ факторов, влияющих на разброс, и провести мероприятия по улучшению состояния процесса (см. рисунок 2.11в).

4. В случае, когда $1,00 > P_p \geq 0,67$, ширина интервала между нижней и верхней границами нормы всего лишь в 4-6 раз превышает стандартное отклонение σ . Когда показатель P_p приближается к 0,67, вероятность появления брака составляет

4,56%. Это означает, что контроль процесса не удовлетворителен. Необходимо наладить строгий контроль процесса и провести сплошной контроль выпускаемых изделий с целью недопущения брака. Вместе с тем нужно провести немедленно исследование факторов, влияющих на разброс, и принять меры к улучшению состояния процесса.

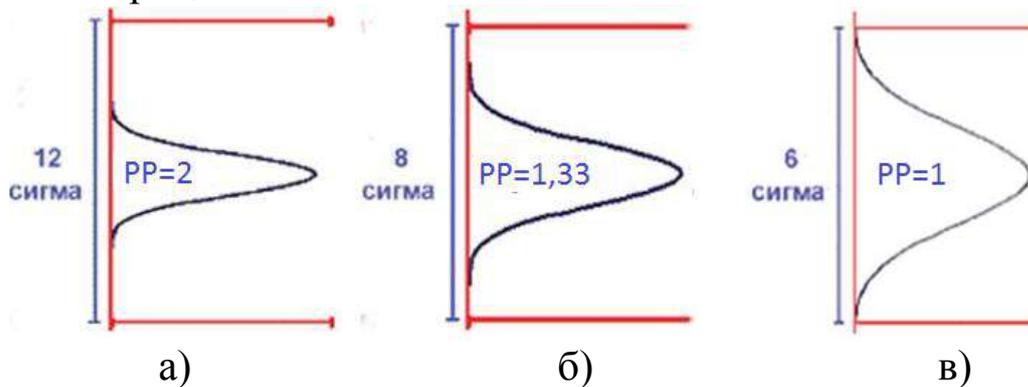


Рисунок 2.11 – Возможные ситуации с точки зрения пригодности процесса

5. В случае, когда $0,67 > P_p$, ширина интервала между нижней и верхней границами нормы не превышает 4σ . Процент брака превышает 4,56%. О таком процессе можно сказать, что он неконтролируем. Необходимо провести сплошной контроль продукции, чтобы предотвратить выпуск бракованных изделий, и одновременно принять меры к повышению качества, выявить причины появления брака. Иногда приходится заново проводить изучение потребностей потребителей, а также пересматривать нормы.

6. Когда $P_p = P_{pk}$ процесс является центрованным, т.е. среднее процесса совпадает с серединой поля допуска, в том случае если целевое значение является серединой интервала, заданного в спецификации. Требуется настройка оборудования.

Анализ точности процесса

После того как были выяснены форма и широта распределения на основании сопоставления с допуском, исследуют, возможно ли по данному технологическому процессу производить качественные изделия. Другими словами, появляется возможность по результатам обследования количественно оценить точность технологических процессов.

Под точностью технологического процесса понимается его способность обеспечивать близость действительных значений контролируемого показателя качества к нормативному (номинальному) значению.

Часто используется обратная величина индекса пригодности, называемая коэффициентом точности процесса:

$$PR = \frac{1}{P_p} = \frac{6s_t}{U - L} \quad (2.14)$$

Точность технологического процесса оценивают исходя из следующих критериев:

$PR \leq 0,75$ технологический процесс точный, удовлетворительный (см. рисунок 2.12а);

$PR = 0,76 - 0,98$ – требует внимательного наблюдения (см. рисунок 2.12б);

$PR > 0,98$ – неудовлетворительный. В этом случае необходимо немедленно выяснить причину появления дефектных изделий и принять меры управляющего воздействия (см. рисунок 2.12в).

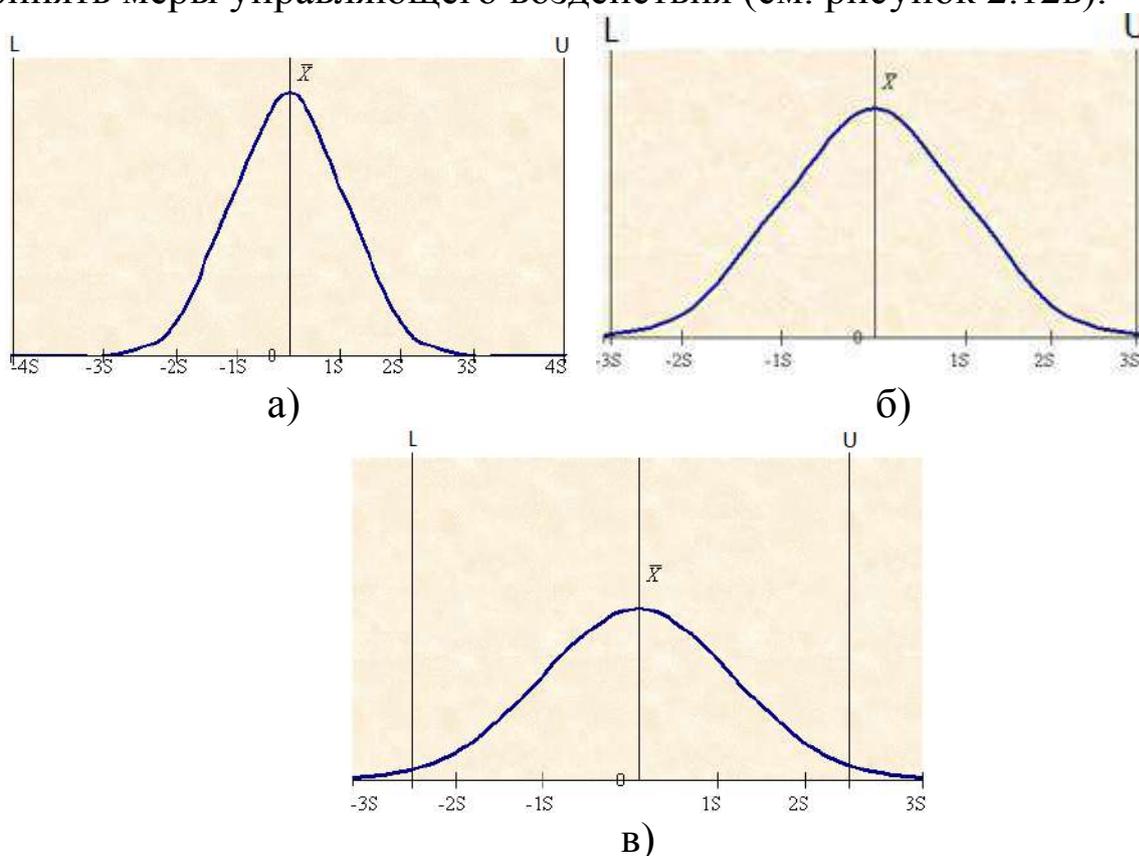


Рисунок 2.12 – Возможные ситуации с точки зрения точности процесса: а) $PR \leq 0,75$; б) $PR = 0,98$; в) $PR > 1$

Исследование распределения статистических данных в программе STATISTICA

Запускаем программу Statistica

На панели задач рабочего стола Windows необходимо последовательно выбрать следующие пункты: ПУСК → Все программы → STATISTICA 10.0 → STATISTICA. При этом загрузится программа STATISTICA (см. рисунок 2.13).

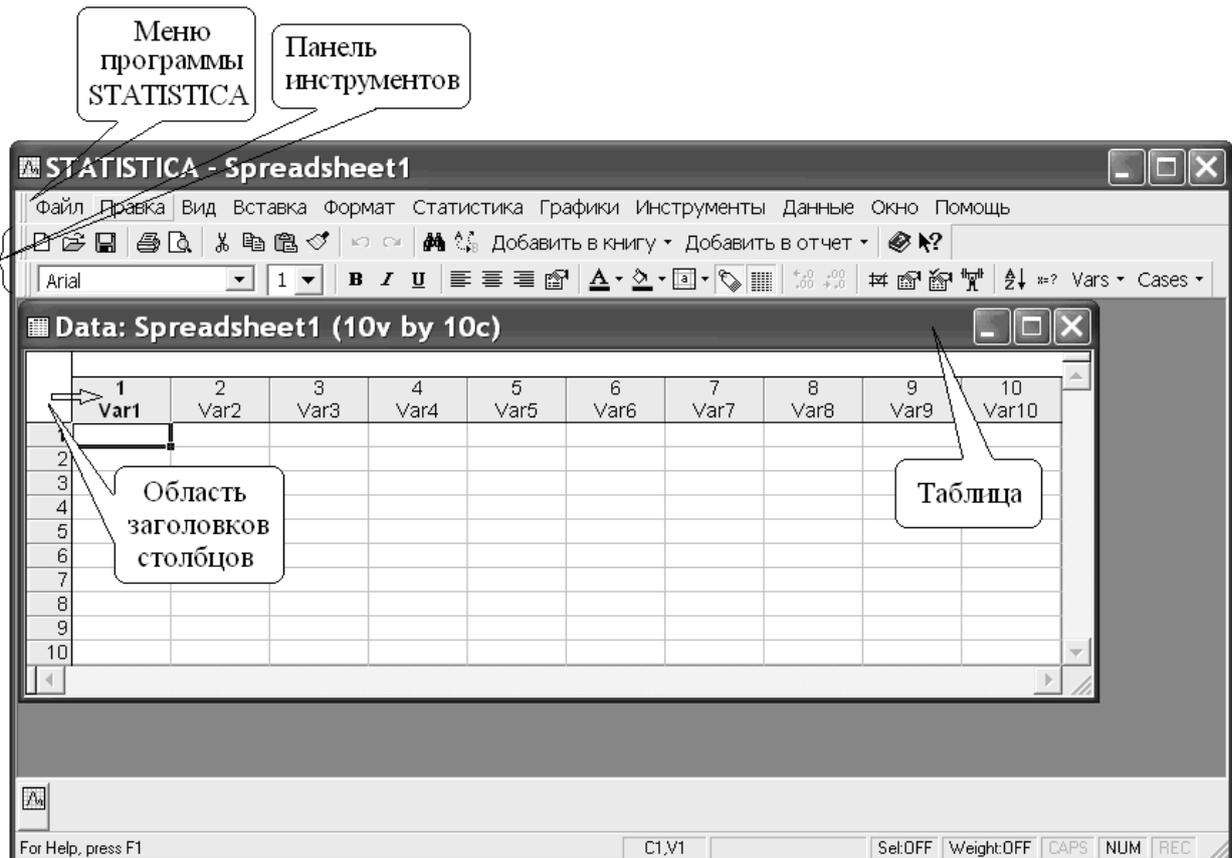


Рисунок 2.13 – Рабочее окно программы STATISTICA

Создаем новую таблицу

Для этого в меню программы STATISTICA выполняем следующие пункты: Файл → Новый...].

Задаём параметры новой таблицы (см. рисунок 2.14), которая будет состоять из одного столбца (число переменных) и 100 строк (число регистров).

Изменим заголовок столбца на «Диаметры осей», по умолчанию он сейчас называется «Var1».

Для этого необходимо открыть окно описания переменных (рисунок 2.15), которое вызывается путем наведения указателя

мыши на область заголовков таблицы (см. рисунок 2.13) и двойного нажатия левой клавиши мыши. В окне описания переменных задаём столбцу имя «Диаметр осей», тип данных «Number», число десятичных знаков «3». Описание переменных завершаем нажатия кнопки «ОК».

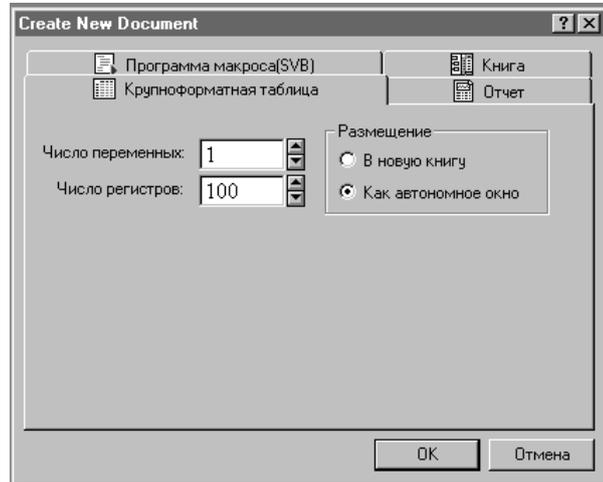


Рисунок 2.14 – Окно новой таблицы

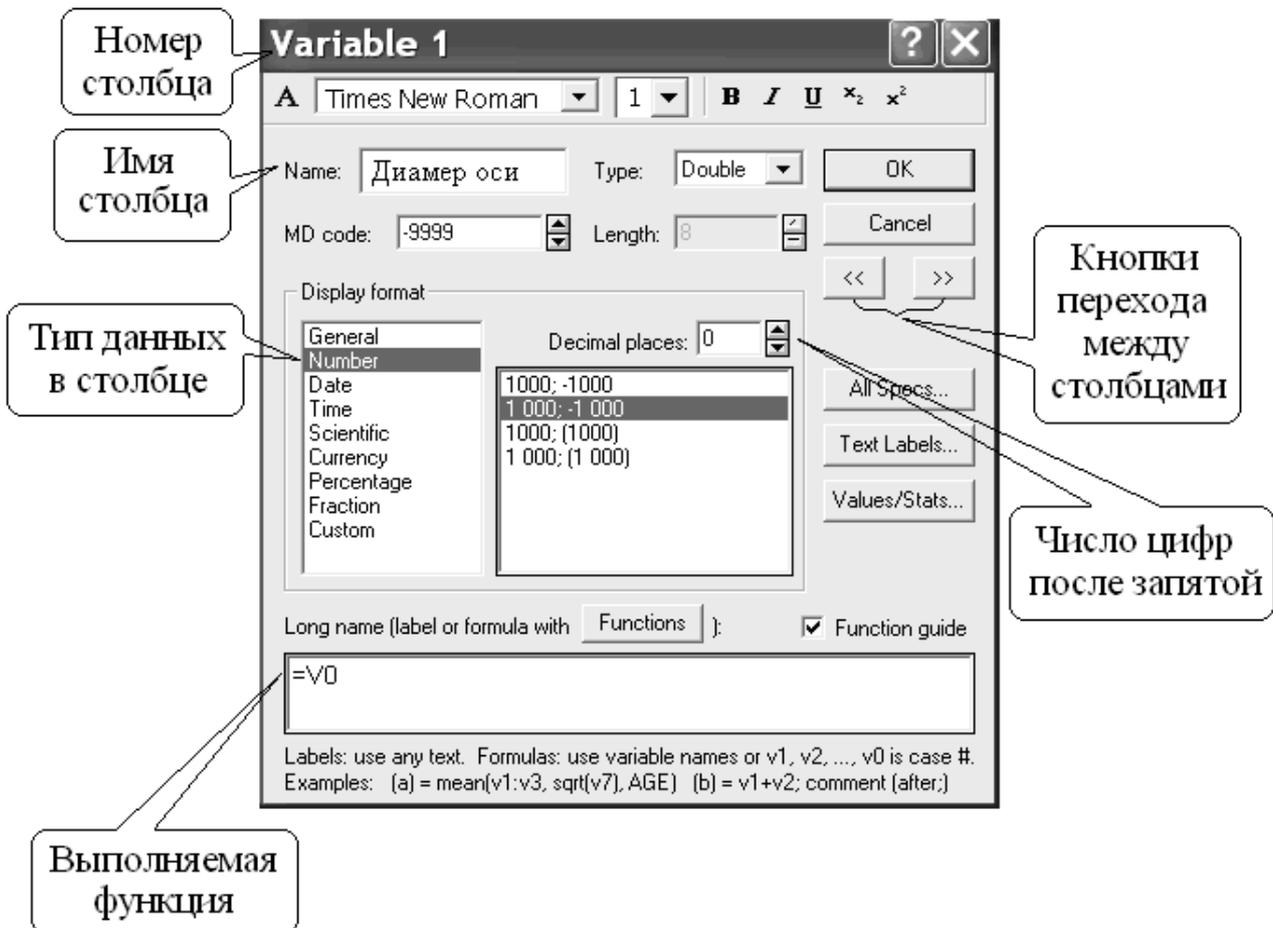


Рисунок 2.15 – Окно описания переменных

Заполняем столбцы переменных данными наблюдений (рисунок 2.16).

Проверка нормальности распределения в программе STATISTICA с помощью построения статистическая функция распределения

Для построения статистической функции распределения открываем следующие вкладки: **Графики** → **2D графики** → **Нормальные графики вероятности** (рисунок 2.17). В

появившемся окне (рис. 2.17) задаём: «Тип графика» - «Normal», «Переменные» - указываем соответствующий столбец с данными. Далее нажимаем кнопку «ОК» и программа строит график (рисунок 2.18).

Data: S...	
1	
Диаметр оси	
1	15,869
2	15,73
3	15,869
4	15,734
5	15,869
6	15,645
7	15,7
8	15,718
9	15,747
10	15,732
11	15,711
12	15,745
13	15,665

Рисунок 2.16 – Таблица

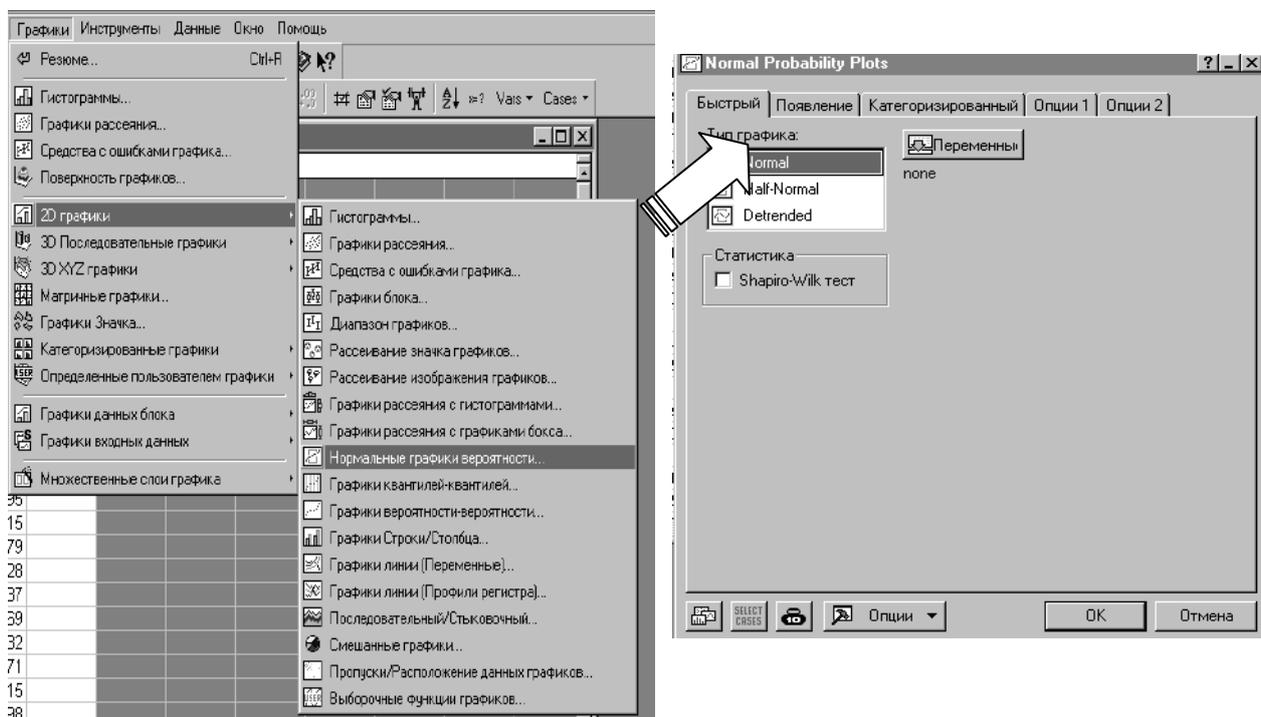


Рисунок 2.18 – Вкладка построения статистической функции распределения

Из построенного графика видно (см. рисунок 2.18), что точки достаточно близко расположены к прямой, и поэтому

распределение можно считать приближенно нормальным. Исключение составляют лишь несколько точек в начале и в конце координат, но это отклонение можно признать не значительным.

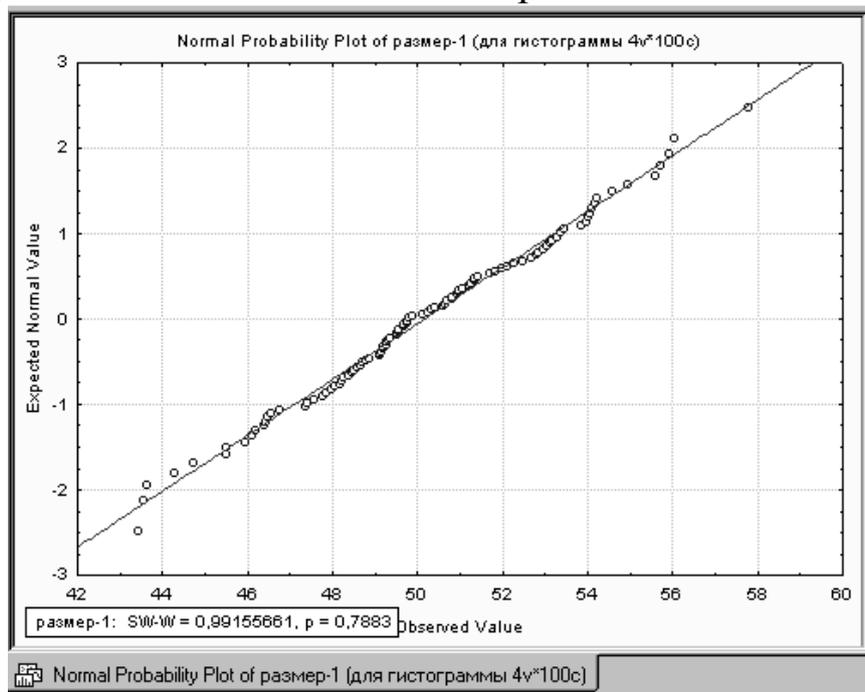


Рисунок 2.18 – Статистическая функция распределения

Проверка нормальности распределения результатов измерения с помощью критерия χ^2 в программе STATISTICA

Для этого выбираем вкладку основного меню **Статистика** → **Настройка распределения**.

В открывшемся окне «Distribution Fitting» выбираем опцию «Continuous Distribution» (рисунок 2.19), в ячейке ниже устанавливаем «Normal» (проверка на сходство с нормальным законом) и далее «ОК». Производим настройку параметров расчета критерия χ^2 в окне настройки «Fitting Continuous Distributions» (рисунок 2.20).

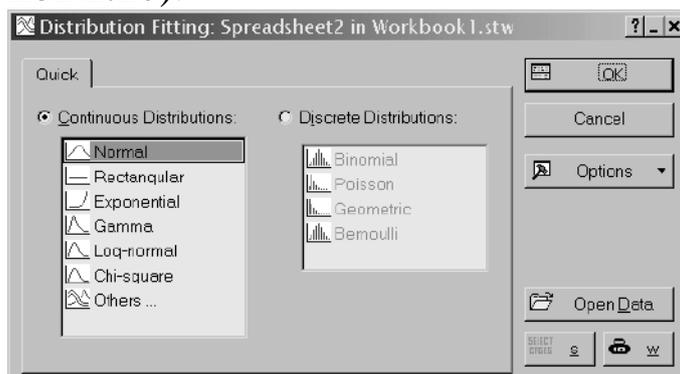
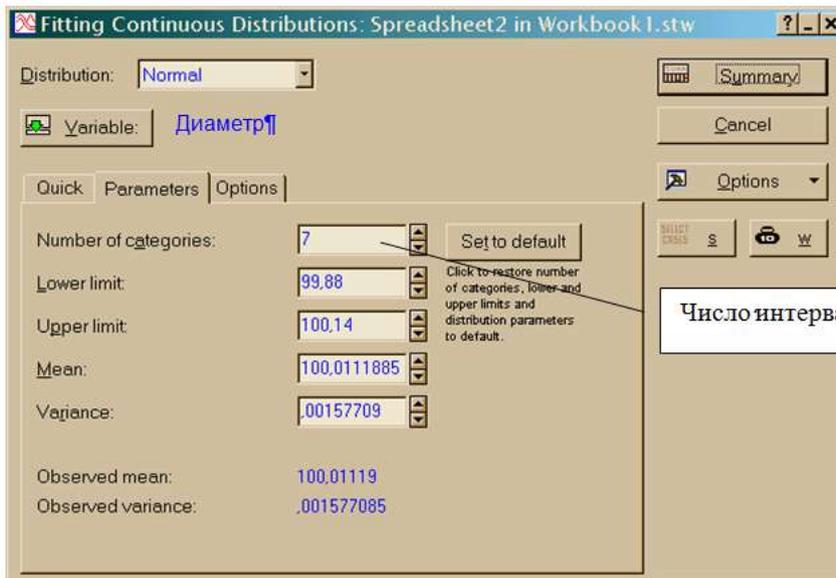
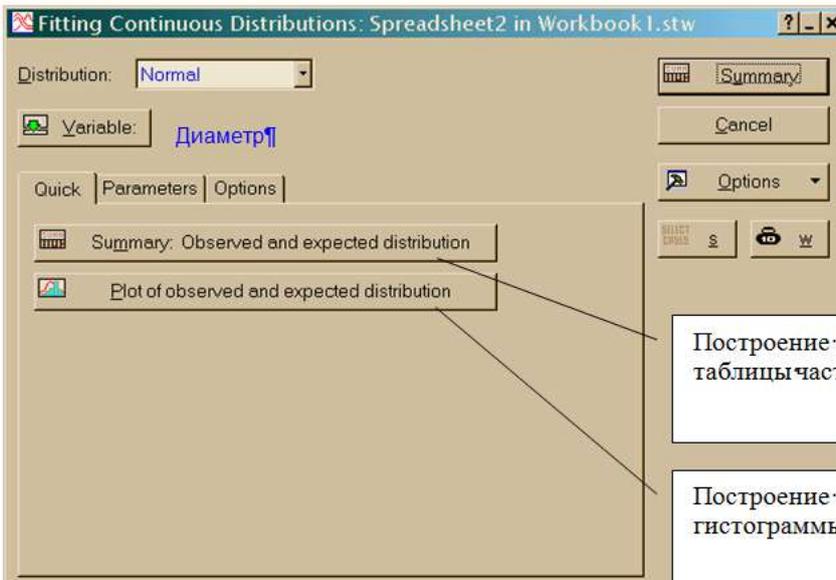


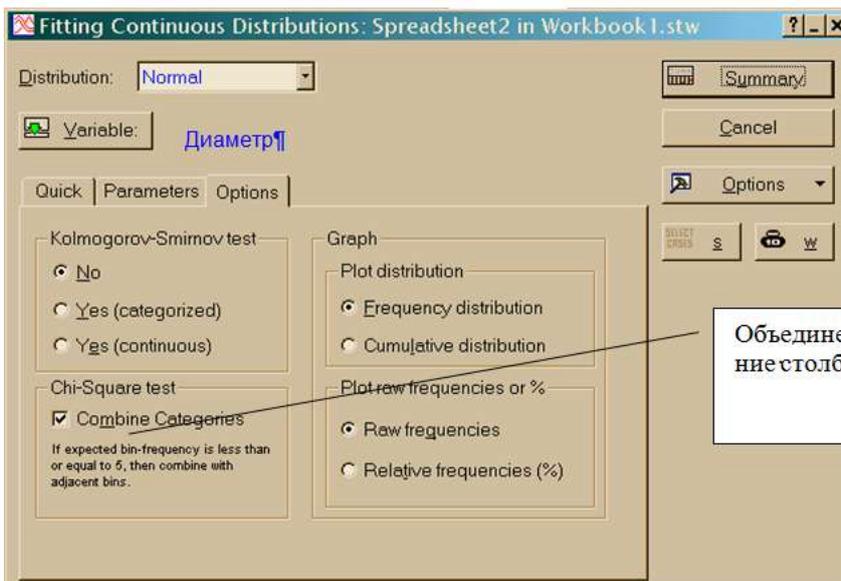
Рисунок 2.19 – Выбор закона распределения



а)



б)



в)

Рисунок 2.20 –
Настройка параметров расчета критерия согласия

Во вкладке «Options» (рисунок 2.20в) на панели «Kolmogorov-Smirnov test» установить опцию «No», на панели «Chi-square test» установить опцию «Combine Categories» (опция объединения столбца с числом попаданий менее 5 с соседним).

Во вкладке «Parameters» (рисунок 2.20а) в ячейке «Number of Categories» (число интервалов) поставить число 7.

Строим таблицу опытных и теоретических частот, во вкладке «Quick» (рисунок 2.20, б) нажимаем кнопку «Summary: Observed and expected distribution». В верхней части таблицы (рисунок 2.21) наблюдаем значения $\chi^2=6,36266$ (Chi-square test), число степеней свободы $k = 4$ (df), вероятность ошибки соответствия нормальному закону $P=0,17365$ (p). Все столбцы данной таблицы можно условно разделить на те, которые характеризуют частоты, относятся к опытным данным и те, которые характеризуют частоты, относящиеся к рассматриваемому закону распределения «Expected».

Upper Boundary	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumul. % Expected	Observed-Expected
<= 99,71429	1	1	1,00000	1,0000	0,28885	0,2889	0,28885	0,2889	0,71115
99,82857	4	5	4,00000	5,0000	4,62964	4,9185	4,62964	4,9185	-0,62964
99,94286	22	27	22,00000	27,0000	24,35399	29,2725	24,35399	29,2725	-2,35399
100,05714	48	75	48,00000	75,0000	42,01916	71,2916	42,01916	71,2916	5,98084
100,17143	22	97	22,00000	97,0000	23,95538	95,2470	23,95538	95,2470	-1,95538
100,28571	2	99	2,00000	99,0000	4,47838	99,7254	4,47838	99,7254	-2,47838
< Infinity	1	100	1,00000	100,0000	0,27460	100,0000	0,27460	100,0000	0,72540

Рисунок 2.21 – Таблица опытных и теоретических частот

Строим гистограмму, во вкладке «Quick» (рисунок 2.20б) нажимаем кнопку «Plot of observed and expected distribution» (рисунок 2.22).

Задаемся уровнем значимости $q = 0,05$, находим по справочным таблицам (см. Приложение Д) значения $\chi_{k;q}^2$ и $\chi_{k;1-q}^2$: $\chi_{4;0,05}^2 = 0,71072$, $\chi_{4;0,95}^2 = 9,48773$, а $\chi_{расч.}^2 = 6,36266$ (рисунок 2.22).

Т.к. $\chi_{4;0,05}^2 < \chi_{расч.}^2 \leq \chi_{4;0,95}^2$ ($0,71072 < 6,36266 < 9,48773$), то распределение результатов наблюдений считают нормальным с вероятностью 95%.

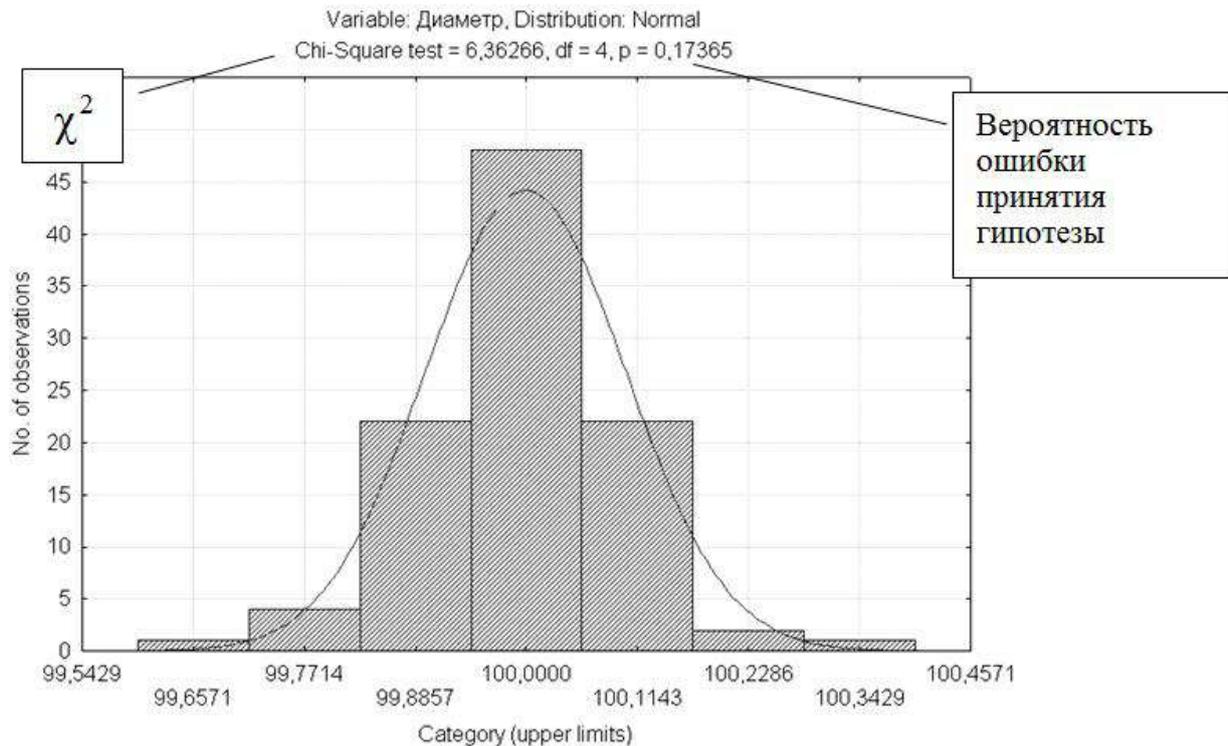


Рисунок 2.22 – Результаты расчета критерия χ^2

Построение гистограммы

Открываем следующую вкладку: **Графики** → **Гистограмма**. В появившемся окне (рисунок 2.23) должны быть установлены опции в панелях «тип графика» – «Regular» (простой); «Распределение» - «Normal» (нормальное); «Тип отображения» - «Standard» (стандартная); «Интервалы» - «Integer mode» (целые числа) и «Auto», а в закладке «Переменные» на панели «Select Variables for Histogram» выбираем номер столбца данных для построения гистограммы и нажимаем кнопку «ОК».

После нажатия кнопки «ОК» в окне «2М Гистограммы» программа строит гистограмму, аналогичную представленной на рисунке 2.24.

Расчет основных характеристик распределения

Для их расчета в программе STATISTICA выполняем следующие действия **Статистика** → **Основная статистика**. В появившемся окне нажимаем на строку Descriptive Statistic (рисунок 2.25).

В появившемся окне (рисунок 2.26) нажимаем на кнопку Variables. В следующем окне задаём переменные для вычислений «Диаметр оси» (рисунок 2.27).

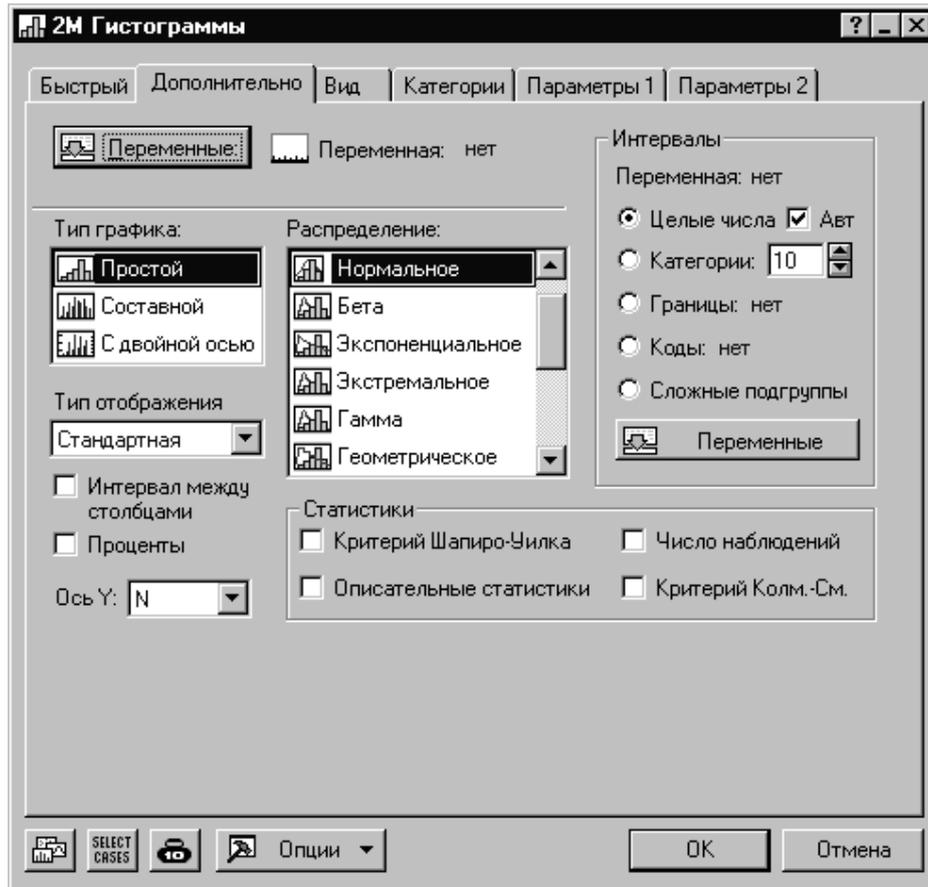


Рисунок 2.23 – Окно построения гистограммы

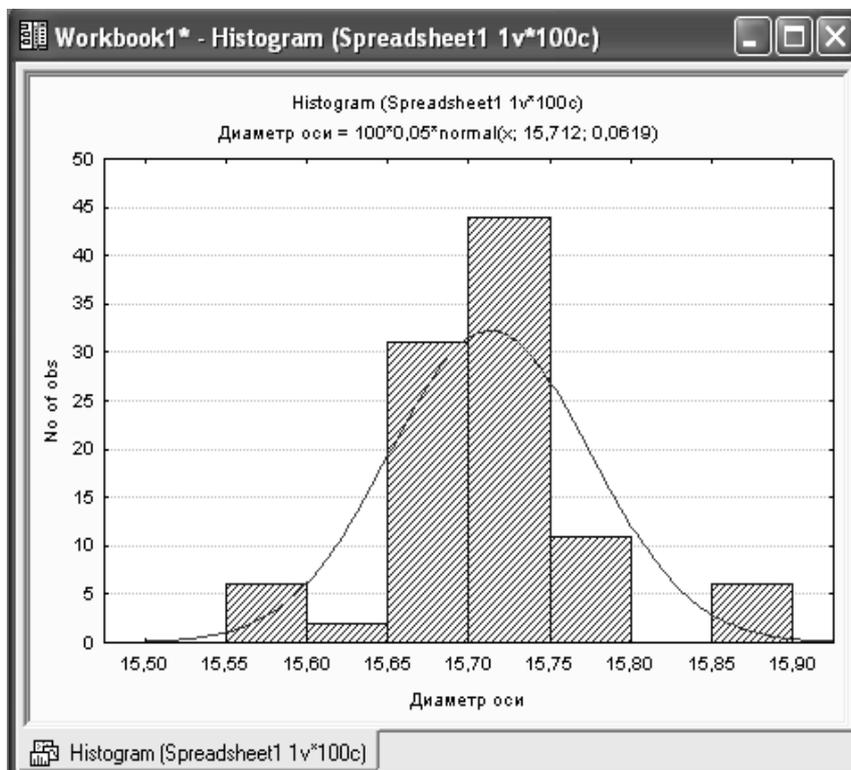


Рисунок 2.24 – Гистограмма

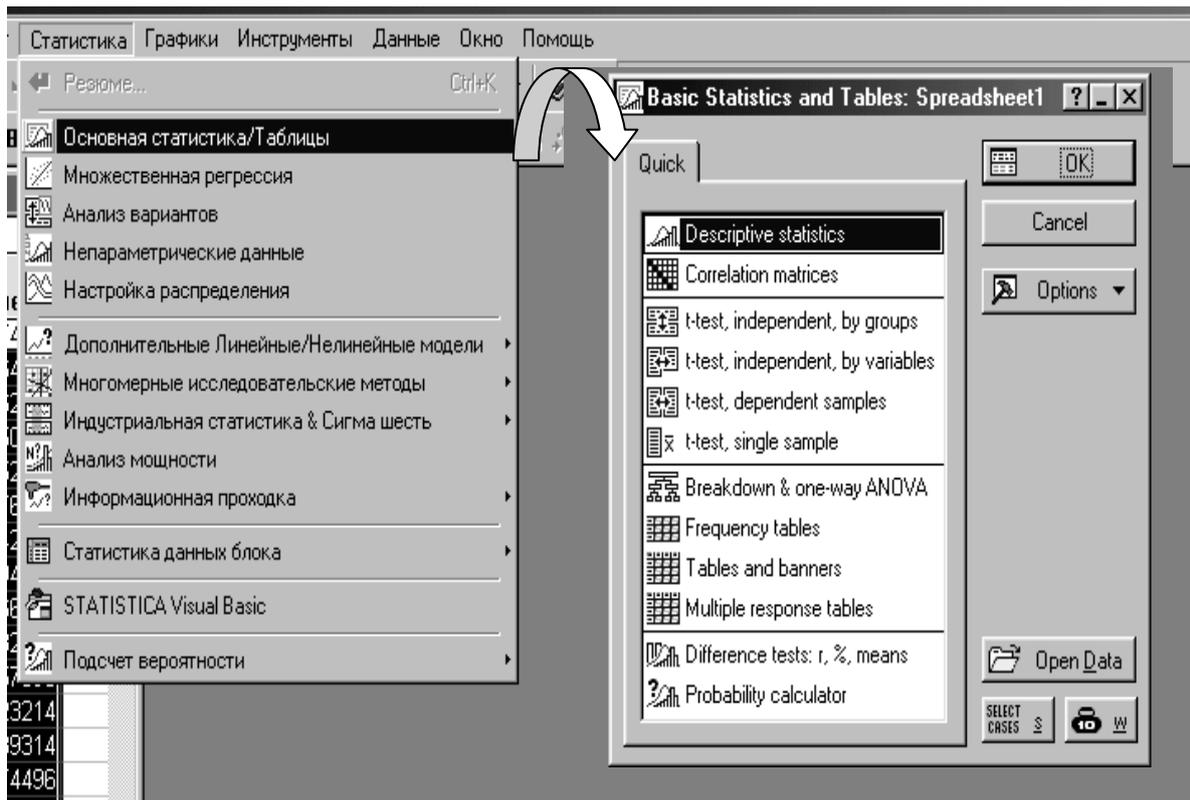


Рисунок 2.25 – Вкладка получения характеристик

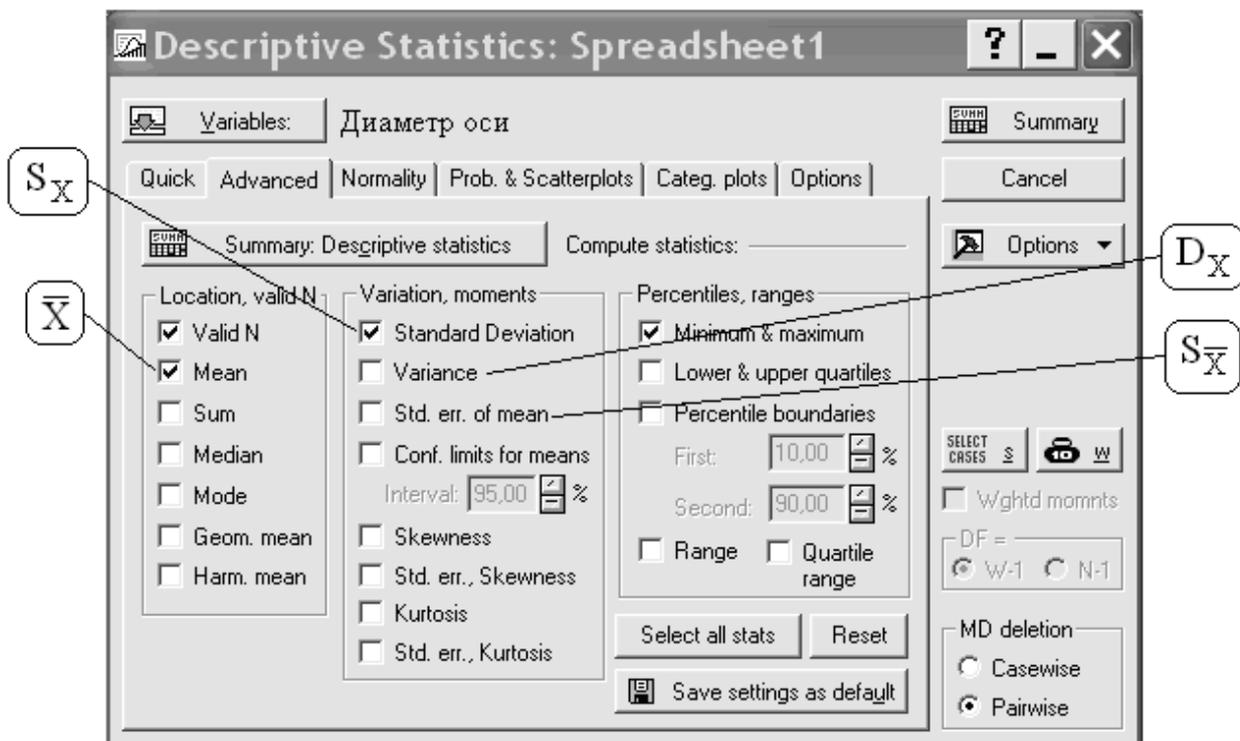


Рисунок 2.26 – Окно настроек основных статистик

В окне настроек (рисунок 2.26) на закладке «Advanced» в панели «Location, valid N» должны быть установлены опции «Valid

N» (число наблюдений), «Mean» (среднее арифметическое \bar{X}), а в панели «Variation, moments» опции «Standard Deviation» (среднее квадратическое отклонение s_x), «Variance» (Дисперсия), «Std. err. of mean» (среднее квадратическое отклонение среднего арифметического $s_{\bar{x}}$). Все остальные опции на данной закладке отключаем и строим таблицу параметров распределения, кнопка «Summary» (рисунок 2.28).

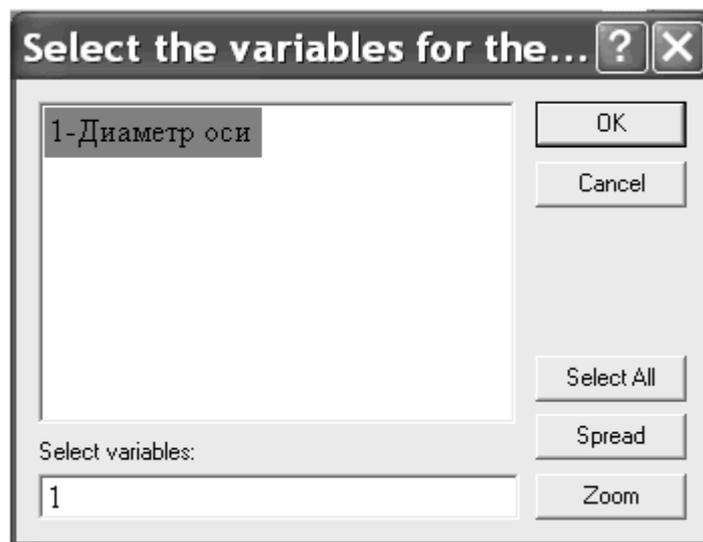


Рисунок 2.27 – Окно выбора переменных

Variable	Valid N	Mean	Variance	Std.Dev.	Standard Error
Диаметр оси	100	15,71196	0,003836	0,061933	0,006193

Рисунок 2.28 – Таблица параметров

Оценка точности и пригодности процесса

Для оценки возможностей процесса в программе STATISTICA выполняем следующие действия Статистика → Индустриальная статистика & Сигма шесть → Анализ процесса. В появившемся окне выбираем строку Process capability analysis & tolerance intervals, raw data. И во вкладке Raw data нажимаем на кнопку Variables. В следующем окне задаем переменные для вычислений «Диаметр оси». В той же вкладке нажимаем кнопку Process specs и в появившемся окне задаем параметры поля допуска (рисунок 2.29).

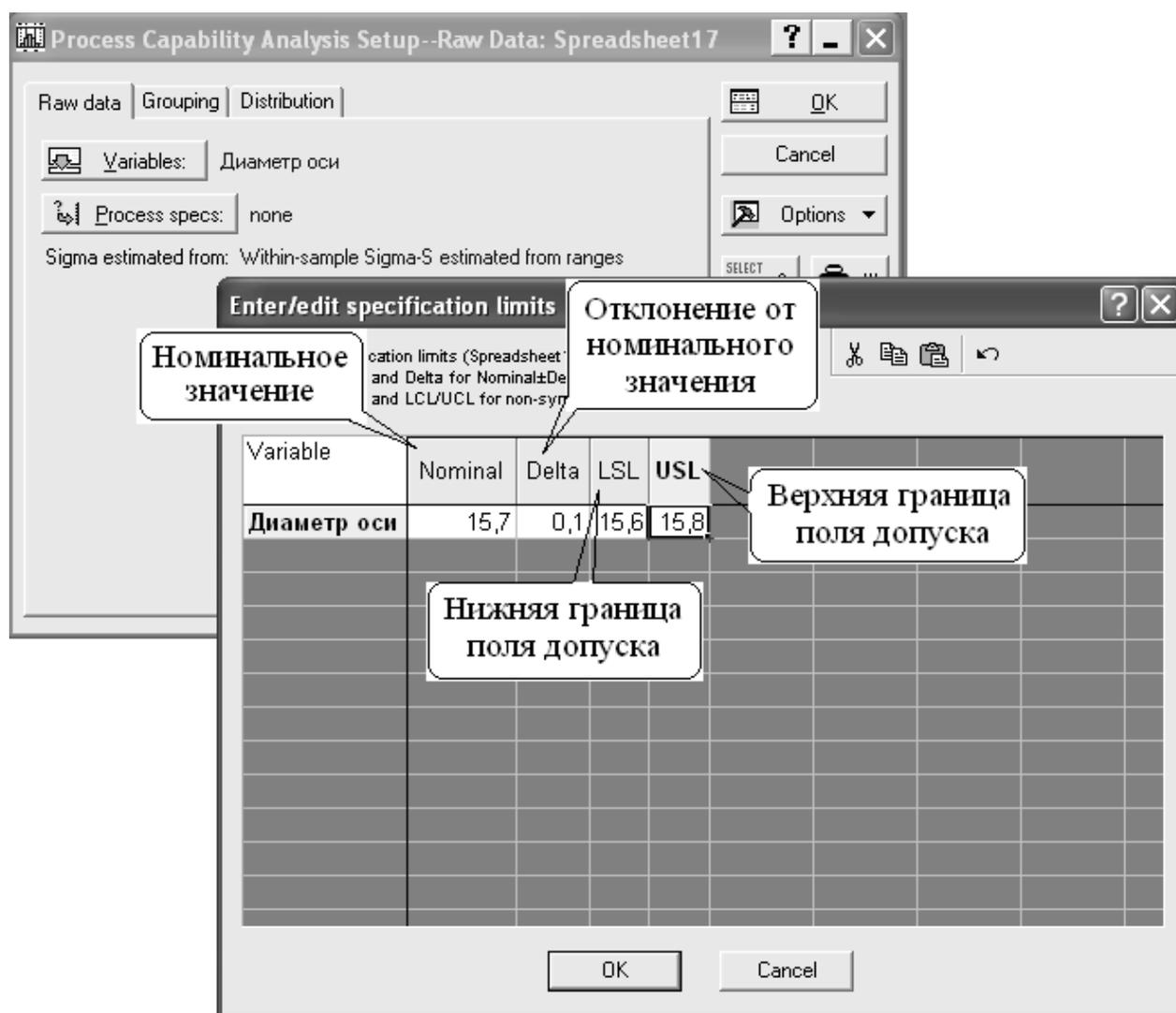


Рисунок 2.29 – Параметры поля допуска

Во вкладке «Grouping» (рисунок 2.30) в ячейке «Constant sample size» (количество измерений) поставить число 100 и в ячейке «Estimate sigma from...» выбрать Variances. Все остальные опции на данной вкладке должны быть отключены. И строим таблицу индексов пригодности процесса, кнопка «Summary» во вкладке соответствующей нормальному или негауссовскому распределению (по результатам проверки нормальности распределения) (рисунок 2.31). Следует обратить внимание, что программа рассчитывает сразу и индексы пригодности, и индексы воспроизводимости. Индексы воспроизводимости используются для оценки возможностей процесса в том случае, если с использованием контрольных карт подтверждена статистическая управляемость процесса.

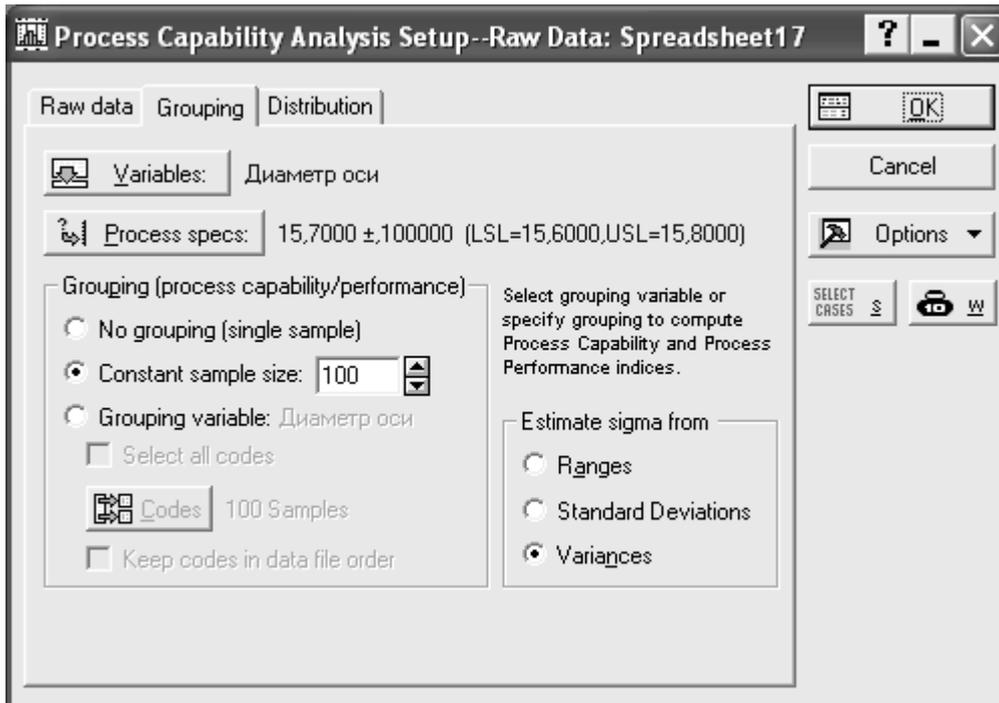


Рисунок 2.30 – Параметры анализа процесса

Variable: Диаметр оси (Sprea...)	
-3,000 *Sigma=15,5262	
+3,000 *Sigma=15,8978	
	Value
Performance Index	
Lower Specification Limit	15,60000
Nominal Specification	15,70000
Upper Specification Limit	15,80000
PP (performance index)	0,53822
PR (performance ratio)	1,85797
PPK (perf. demonstr. excell.)	0,47385
PPL (lower performance index)	0,60259
PPU (upper performance index)	0,47385

Рисунок 2.31 – Индексы пригодности процесса

Нажав кнопку «Summary histogram» программа строит гистограмму с обозначение границ поля допуска, номинального значения и контрольных границ регулирования процесса на уровне $\pm 3\sigma$ (рисунок 2.32).

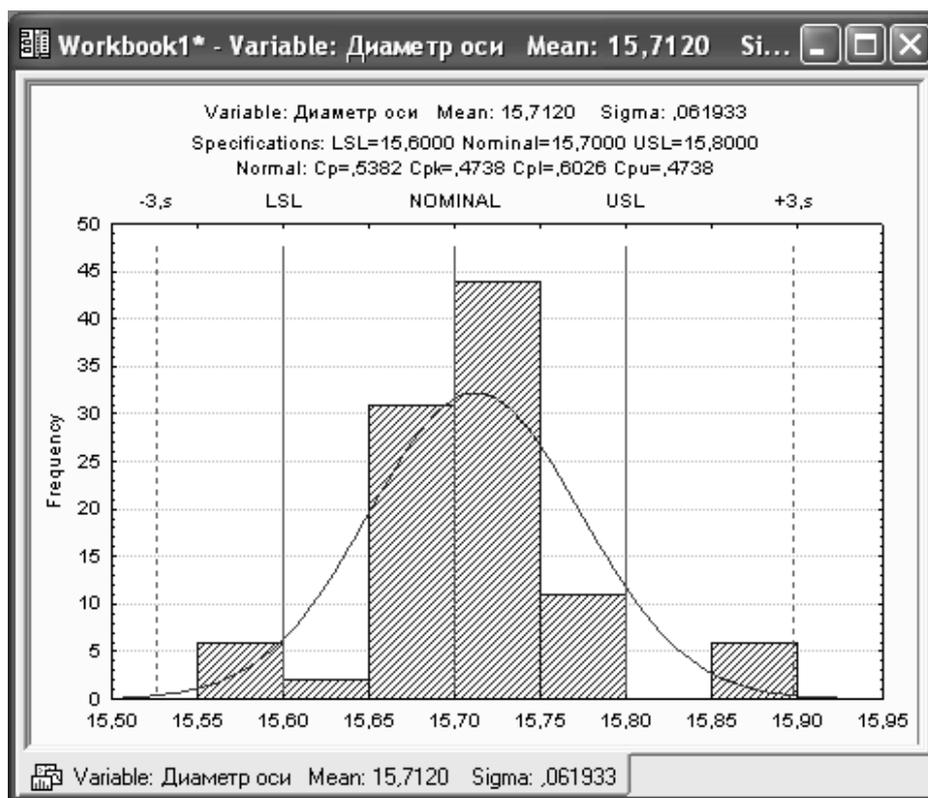


Рисунок 2.32 – Расположение гистограммы в пределах поля допуска

Построение гистограммы в программе LibreOffice

Для определения характера рассеяния показателя строим гистограмму.

Порядок построения гистограммы:

1. *Намечаем исследуемый показатель качества.* В данном случае это коэффициент деформации материала.

2. *Проводим измерения.* Должно быть не менее 30...50 данных, оптимально – около 100.

Результаты измерений коэффициента деформации представлены в таблице 2.1.

Результаты измерений вводим в электронную таблицу. В ячейку A1 вводим заголовок работы. Начиная с ячейки A3 вводим в столбец порядковые номера измерений с 1 по 100, например при помощи команды **Правка** ▶ **Заполнить** ▶ **Прогрессия...** . В ячейки B3:B102 вводим значения коэффициента деформации из таблицы 2.1.

3. *Вводим единицу измерений.* Единица измерений равна точности, с которой проводились измерения, в данном случае 0,1. Вводим единицу измерений в ячейку E2.

Таблица 2.1

Результаты измерений коэффициента деформации

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

4. *Находим минимальное и максимальное значения выборки.* Минимальное и максимальное значения выборки находим с помощью статистических функций МИН и МАКС соответственно в ячейках E3 и E4. При этом интервал для этих функций указываем от ячейки B3 до ячейки B102.

5. *Находим размах выборки* в ячейке E5 как разность между максимальным и минимальным значениями выборки.

6. *Определяем предварительное количество интервалов $K_{предв}$ как квадратный корень из объёма выборки N .* Количество интервалов находим в ячейке E6. Поскольку количество интервалов должно быть целым числом, т.е. полученный квадратный корень следует округлить до целого значения, то сначала в ячейку E6 вводим математическую функцию ОКРУГЛ. В строке **Количество_цифр** этой функции указываем 0, т.к. необходимо округление до целого числа. Затем переводим курсор в строку **Число** и в качестве аргумента функции ОКРУГЛ встраиваем функцию КОРЕНЬ. Для этого в строке формул открываем список функций, выбираем **Другие функции...** и открываем математическую функцию КОРЕНЬ. В качестве аргумента функции КОРЕНЬ опять при помощи списка в строке формул выбираем статистическую функцию СЧЁТ, в качестве аргумента которой вводим диапазон ячеек от B3 до B102. Поскольку функция СЧЁТ подсчитывает количество чисел в указанном диапазоне, т.е. в данном случае объём выборки, то будет получено значение 100. Затем функция КОРЕНЬ пересчитает это значение в 10, а функция

ОКРУГЛ округлит его до целых, т.е. до 10. В целом формула в ячейке E6 будет выглядеть примерно так:
 $=\text{ОКРУГЛ}(\text{КОРЕНЬ}(\text{СЧЁТ}(\text{В3:В102})));0)$

7. *Определяем ширину интервала* в ячейке E7 по формуле $h = R/K_{\text{предв}}$ с округлением до единицы измерения, т.е. в нашем случае до десятых долей. Формула в ячейке E7 будет выглядеть так:
 $=\text{ОКРУГЛ}(E5/E6;1)$.

8. *Вводим номера интервалов.* Для этого в ячейку D9 вводим заголовок столбца **№ инт.** Начиная с ячейки D10 вводим номера интервалов с 1 примерно до 25.

9. *Рассчитываем границы и середины интервалов.* В ячейке E10 рассчитываем нижнюю границу первого интервала по формуле
 $X_{\min} - \text{ед.изм.}/2$

Для этого в ячейку E10 вводим формулу $=E3-E2/2$ и получаем значение нижней границы первого интервала 0,05.

В ячейке E11 рассчитываем нижнюю границу второго интервала, прибавляя к нижней границе первого интервала значение шага. Формула в ячейке E11 будет выглядеть $=E10+E7$. После указания необходимой абсолютной адресации копирует эту формулу в диапазон E12:E34.

В ячейке F10 рассчитываем верхнюю границу первого интервала, прибавляя к его нижней границе значение шага. После указания необходимой абсолютной1 адресации полученную формулу копируем в диапазон F11:F34.

В ячейке G10 рассчитываем среднее значение первого интервала, например, по статистической формуле СРЗНАЧ. Полученную формулу копируем в диапазон G11:G34.

Поскольку уже в десятом интервале нижняя граница равна 1,85. что больше X_{\max} , то необходимое количество интервалов равно 9. Поэтому содержимое ячеек диапазона D19:F34 следует очистить.

10. *Подсчитываем частоты появления результатов измерений в интервалах.* В ячейке H10 рассчитываем частоту для первого интервала при помощи статистической функции СЧЁТЕСЛИ. Функция СЧЁТЕСЛИ подсчитывает количество непустых ячеек в указанном диапазоне, удовлетворяющих заданному условию. Следует подсчитать, сколько раз в диапазоне В3:В102 встречаются ячейки, значения которых находятся в границах первого интервала,

т.е. больше 0,05, но меньше 0,25. Таким образом, надо подсчитать ячейки, значения которых удовлетворяют двойному условию. Однако функция СЧЁТЕСЛИ использует только одинарное условие. Поэтому в формуле, записываемой в ячейке Н10, функцию СЧЁТЕСЛИ используем дважды. Сначала в функции СЧЁТЕСЛИ вводим диапазон В3:В102 и условие ">0,05". (к сожалению, нельзя указать условие ">Е10", ссылаясь на значение нижней границы интервала, поскольку функция СЧЁТЕСЛИ использует условие критерий в форме числа, выражения или текста, но не в форме ссылки на ячейку). Затем переводим курсор в строку формул, ставим знак минус, вновь вводим функцию СЧЁТЕСЛИ, указываем в ней диапазон В3:В102 и условие ">0,25". В результате получаем расчётную формулу $=\text{СЧЁТЕСЛИ}(\text{В3:В102};">0,05")-\text{СЧЁТЕСЛИ}(\text{В3:В102};">0,25")$, по которой рассчитывается частота для первого интервала. После указания абсолютной адресации для интервалов копируем эту формулу в диапазон Н11:Н18. Поскольку в копируемой формуле границы интервалов были указаны численными значениями, то в формулах ячеек диапазона Н11:Н18 следует исправить численные значения границ на соответствующие тому или иному диапазону. Например, в ячейке Н11 формула будет выглядеть так: $=\text{СЧЁТЕСЛИ}(\text{\$B\$3:\$B\$102};">0,25")-\text{СЧЁТЕСЛИ}(\text{\$B\$3:\$B\$102};">0,45")$.

Результаты расчётов показаны на рисунке 2.33.

Н11		=СЧЁТЕСЛИ(\$B\$3:\$B\$102;">0,25")-СЧЁТЕСЛИ(\$B\$3:\$B\$102;						
	А	В	С	Д	Е	Ф	Г	Н
1	Лаб. работа 6. Гистограммы							
2	№	Козф. деформ.	Ед. изм.=	0,1				
3	1	0,9	Хмин =	0,1				
4	2	0,6	Хмах =	1,8				
5	3	0,5	R =	1,7				
6	4	0,6	Кпредв =	10				
7	5	0,7	h =	0,2				
8	6	0,8						
9	7	1	№ инт.	Ниж. гр.	Верх. гр.	Середина	Частота f	
10	8	1,4	1	0,05	0,25	0,15	2	
11	9	1,1	2	0,25	0,45	0,35	8	
12	10	1,5	3	0,45	0,65	0,55	13	
13	11	1,5	4	0,65	0,85	0,75	15	
14	12	0,1	5	0,85	1,05	0,95	20	
15	13	0,8	6	1,05	1,25	1,15	17	

Рисунок 2.33 – Расчёт данных для построения гистограммы

11. *Строим гистограмму распределения.* Открываем мастер диаграмм, выбираем тип **Гистограмма** и вид **Обычная гистограмма** отображает значения различных категорий. На втором шаге на вкладке **Диапазон данных** указываем диапазон H10:H18. На вкладке **Ряд** в строке **Подписи по X** указываем диапазон G10:G18 (возможно указание диапазона E10:F18). На третьем шаге вводим заголовки по осям, а также убираем легенду и линии сетки. После создания диаграммы редактируем её, используя контекстное меню. В частности, открыв контекстное меню на одном из столбцов диаграммы, выбираем команду **Формат рядов данных...**, вкладку **Параметры**, и устанавливаем ширину зазора 0.

Готовая гистограмма показана на рисунке 2.34а. Возможно представление гистограммы в виде непрерывной кривой или ломаной линии. Для этого надо в области гистограммы открыть контекстное меню, выбрать команду **Тип диаграммы...**, выбрать диаграмму **Точечная** и соответствующий её вид (рисунок 2.34 б, в).

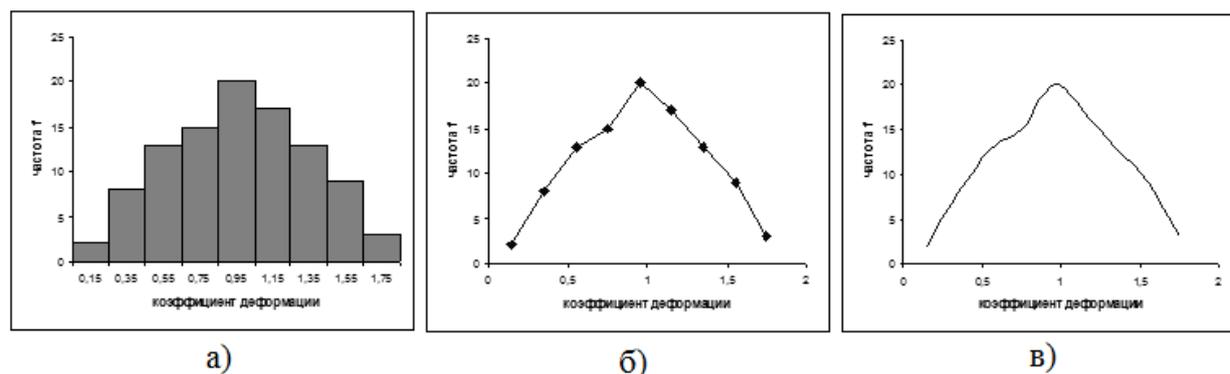


Рисунок 2.34 – Гистограмма в виде столбиковой диаграммы (а), ломаной линии (б) и непрерывной кривой (в)

Полученная гистограмма близка к обычной гистограмме с двусторонней симметрией, что указывает на стабильность процесса.

Порядок выполнения работы

Получив у преподавателя исходные данные для выполнения лабораторной работы, студент изучает теоретические сведения. Далее выполняет на компьютере проверку нормальности распределения; построение гистограммы, расчет основных характеристик полученного распределения и оценку точности и пригодности процесса.

Затем делает выводы о нормальности распределения; визуальную оценку распределения; визуальную оценку расположения гистограммы в пределах поля допуска; анализ пригодности процесса и анализ точности процесса.

Содержание отчета

Отчет по лабораторной работе должен содержать следующие пункты:

- название лабораторной работы;
- цель работы;
- краткие теоретические сведения;
- краткое описание хода выполнения работы;
- индивидуальное задание для выполнения лабораторной работы;
- результаты выполнения работы: таблица с исходными данными; статистическая функция распределения; таблица опытных и теоретических частот; результаты расчета критерия хи-квадрат; гистограмма, таблица с параметрами распределения, расчет индексов пригодности, гистограмма в пределах поля допуска.
- выводы о нормальности распределения, точности и пригодности процесса.

Контрольные вопросы

1. Что такое «гистограмма»?
2. Для чего применяется гистограмма?
3. Для чего необходима гистограмма при проведении контроля показателей качества продукции?
4. Что характеризуют различные типы гистограмм?
5. Что характеризует основные параметры распределения?
6. Что характеризует индексы пригодности?
7. Охарактеризуйте расположение статистических данных относительно поля допуска.

Лабораторная работа №3

Выявление основных причин появления проблемы на основании анализа диаграммы Парето

Цель работы: получить сведения о назначении, видах диаграммы Парето и получить практические навыки обработки статистических данных, построении и анализе диаграммы Парето в программе *STATISTICA* или *LibreOffice*.

Задание для самостоятельного выполнения.

Рабочие А и В работают на одном станке. Рабочий А – в первую смену, а рабочий В – во вторую. Проанализируйте данные выданные преподавателем, составляя различные диаграммы Парето.

Краткие теоретические сведения

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

С помощью диаграммы Парето можно по всем видам брака для конкретной продукции объективно и полно оценить потери предприятия и установить важность тех или иных факторов в показателях качества продукции.

Чрезвычайно важно прояснить картину распределения потерь. И благодаря диаграмме Парето можно сначала сосредоточить внимание на устранении дефектов, которые приводят к наибольшим потерям.

Виды диаграмм Парето

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, чтений, 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 в программе STATISTICA

Создадим новую таблицу

Для построения диаграммы Парето создаём новую таблицу, состоящий из 2 столбцов и 6 строк. В первый столбец вводим название типов дефектов (Дефект 1, Дефект 2 и т.д.), а во второй столбец число появлений данного дефекта ходе анализа (рисунок 3.1).

	1	2
	Типы дефектов	Число дефектов
1	Дефект 1	5
2	Дефект 2	10
3	Дефект 3	8
4	Дефект 4	15
5	Дефект 5	6
6	Прочие	4

Рисунок 3.1 – Таблица с данными

Открываем следующую вкладку: Статистика → Индустриальная статистика & Сигма шесть → Качество диаграммы управления (рисунок 3.2). В появившемся окне выбираем строку «Pareto chart analysis» (рисунок 3.3).

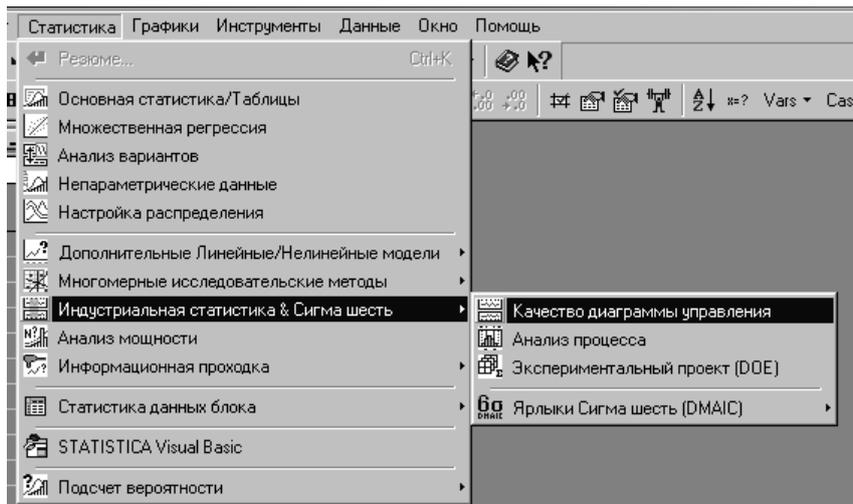


Рисунок 3.2 – Вкладка построения диаграммы

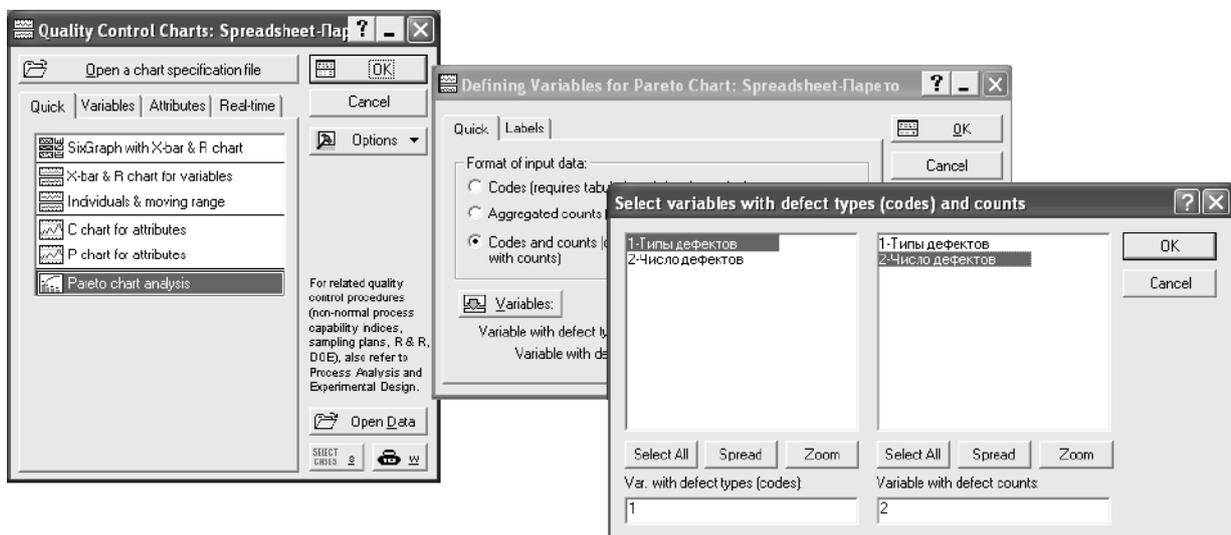


Рисунок 3.3 – Описание переменных для построения диаграммы Парето

Строим диаграмму

В окне «Defining Variables for Pareto Chart» (рисунок 3.3) на панели «Format of input data:» устанавливаем опцию «Codes and counts».

Во вкладке «Variables» указываем две переменных – «Типы дефектов» (в левом окне) и «Число дефектов» (в правом). После нажатия кнопок ОК программа строит диаграмму Парето, аналогичную представленной на рисунке 3.4.

Строим таблицу диаграммы

В окне «Pareto Chart» (рисунок 3.5) нажимаем кнопку «Display chart summary», при появляется таблица данных, по которым была построена диаграмма Парето (рисунок 3.6).

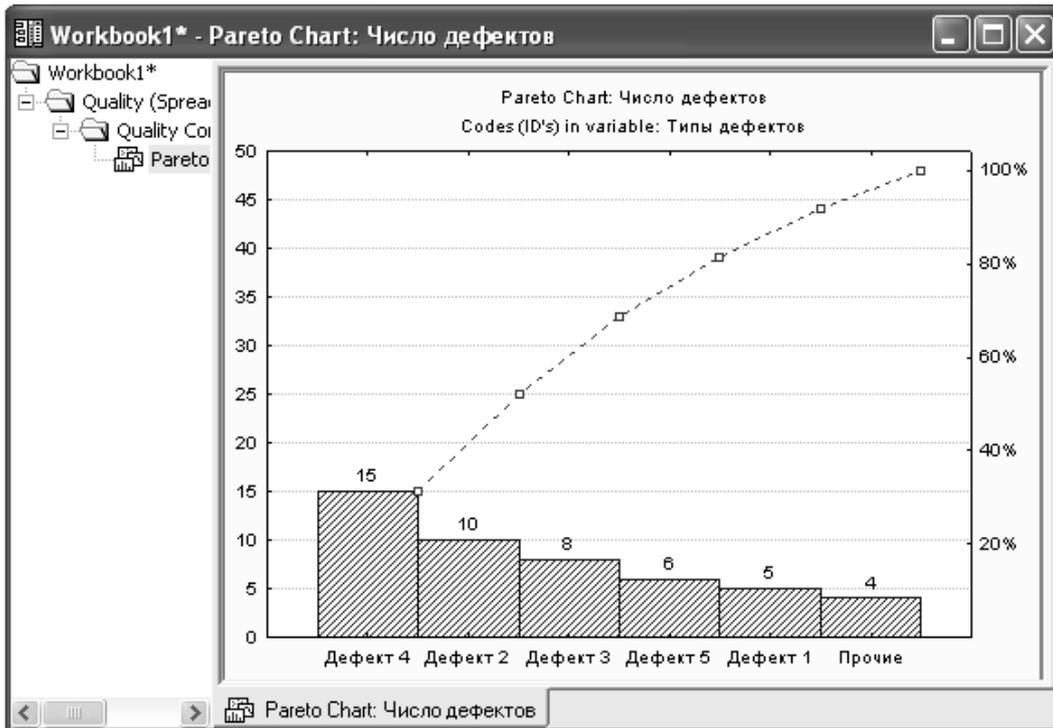


Рисунок 3.4 – Диаграмма Парето



Рисунок 3.5 – Окно построения таблицы данных

Pareto Chart: (Spreadsheet-Парето)				
Codes (ID's) in variable: Типы дефектов				
Category	Value	Cumulative Value	Percent	Cumulative Percent
Дефект 4	15,00000	15,00000	31,25000	31,25000
Дефект 2	10,00000	25,00000	20,83333	52,08333
Дефект 3	8,00000	33,00000	16,66667	68,75000
Дефект 5	6,00000	39,00000	12,50000	81,25000
Дефект 1	5,00000	44,00000	10,41667	91,66667
Прочие	4,00000	48,00000	8,33333	100,00000

Рисунок 3.6 – Таблица диаграммы Парето

Построение диаграммы Парето в программе *LibreOffice*

Исследуем проблему появления брака при выпуске деталей с помощью построения диаграммы Парето.

С учётом того, что потери от брака одной детали каждого вида примерно одинаковы, в качестве единицы измерения выбираем число дефектных деталей каждого вида. После заполнения контрольных листков получаем данные, представленные в таблице 3.1.

Таблица 3.1

№ детали	1	2	3	4	5	6	Прочие
Число дефектных деталей	255	101	59	39	26	15	11

По полученным данным разрабатываем таблицу для проверок данных. Создаём новую книгу *LibreOffice*. В ячейке A1 вводим заголовок работы. В ячейки A3:E3 вводим заголовки: **№ детали**, **Число дефектных деталей**, **Накопленная сумма деталей**, **Процент деталей**, **Накопленный процент**. Для компактного размещения заголовков выделяем третью строку и используем команду **Формат ▶ Ячейки...**, вкладку **Выравнивание**, режим выравнивания по вертикали **По центру**, режим отображения **Переносить по словам**.

В ячейки A4:B10 вводим данные из таблицы 3. В ячейку A11 вводим заголовок **Итого**. В ячейке B11 рассчитываем суммарное число дефектных деталей при помощи математической формулы СУММ.

Для расчёта накопленной суммы деталей в ячейку C4 вводим значение 255, т.е. число дефектных деталей 1. В ячейке C5 суммируем число дефектных деталей 1 и 2, т.е. вводим формулу =C4+B5. Для расчёта накопленной суммы деталей в остальных ячейках копируем формулу из ячейки C5 в диапазон C6:C10.

Для расчёта процента деталей следует делить число дефектных деталей каждого вида на общее число дефектных деталей и умножать на 100. Таким образом, в ячейку D4 вводим формулу =B4/B11*100. После указания необходимой абсолютной адресации копируем эту формулу в диапазон D5:D10. В ячейке D11 рассчитываем суммарный процент, который должен составить 100%.

Для расчёта накопленного процента деталей в ячейку E4 значение (только значение, а не формулу) из ячейки D4. Для этого используем команды **Правка ▶ Копировать** и **Правка ▶ Специальная вставка...**. В ячейке E5 суммируем процент дефектных деталей 1 и 2, т.е. вводим формулу =E4+D5. Для расчёта накопленного процента в остальных ячейках копируем формулу из ячейки E5 в диапазон E6:E10.

По таблице для проверок данных строим диаграмму Парето. Для этого открываем в мастере диаграмм вкладку **Нестандартные**, выбираем диаграмму типа **График/гистограмма 2**. На втором шаге указываем диапазон данных A4:B10; E4:E10. На третьем шаге вводим заголовки и убираем легенду.

После создания диаграммы мастером диаграмм редактируем её при помощи контекстных меню. В частности, максимальное значение шкалы **Число дефектных деталей** указываем 506, а минимальное 0. Максимальное значение шкалы **Накопленный процент** указываем 100. Открываем контекстное меню на одном из столбцов, выбираем команду **Формат рядов данных...**, вкладку **Параметры**, и устанавливаем ширину зазора 0.

Результаты расчётов и построений показаны на рисунке 3.7.



Рисунок 3.7 – Построение диаграммы Парето по числу дефектных деталей

Как видно из диаграммы (см. рисунок 3.7), к группе А можно отнести детали 1 и 2 (70% от брака), к группе В – детали 3,4,5, к группе С – детали 6 и прочие.

Для выяснения наиболее важных дефектов целесообразно построить диаграммы Парето появления дефектности в деталях 1 и 2.

Рассмотрим построение такой диаграммы для детали 1. В качестве единицы измерения выбираем сумму потерь от брака, млн. руб. После исследования явлений дефектности получили данные, представленные в таблице 3.2.

Таблица 3.2

Дефект	Сумма потерь, млн. руб.
Шаг резьбы завышен	1,5
На режущей кромке резца налипы	6,9
Зависание	1,9
Пропуск операции	0,4
Осталась чернота	0,9
Скос кромки увеличен	0,6
Наружный диаметр занижен	8,3
Прочие	0,2

Диаграмма Парето, построенная по этим данным, показана на рисунке 3.8.

Как видно из диаграммы (см. рисунок 3.8), к группе А можно отнести занижение наружного диаметра и налипы на режущей кромке резца (73% от суммы потерь), к группе В – зависание, завышение шага резьбы, остаточную черноту, к группе С – увеличение скоса кромки, пропуск операции и прочие.

Для выяснения наиболее важных причин потерь целесообразно построить диаграммы Парето по причинам занижения наружного диаметра и налипов на режущей кромке резца. При построении такой диаграммы для причин занижения наружного диаметра после заполнения контрольных листков получили данные, представленные в таблице 3.3

Таблица 3.3

Причина	Число дефектов
Смещение копира	53
Неопытность оператора	11
Неточность рабочего инструмента	4
Устаревший чертёж	98
Ошибки в управлении станком	20
Неточность станка	8
Прочие	7

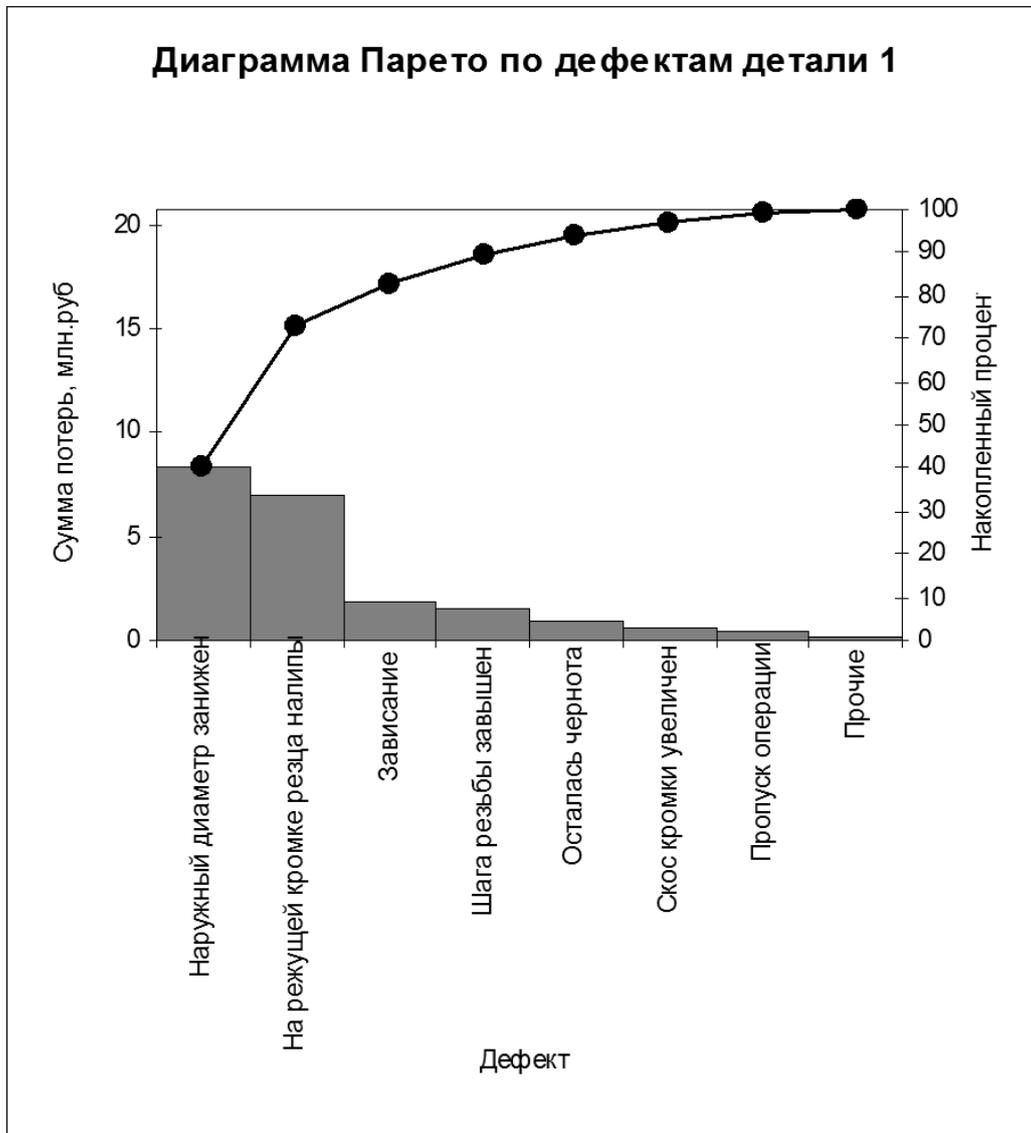


Рисунок 3.8 – Диаграмма Парето по дефектам детали 1

По этим данным необходимо построить диаграмму Парето, выявить причины занижения наружного диаметра группы А и провести по ним корректирующие мероприятия. После этого можно вновь построить диаграмму Парето для изменившихся условий, чтобы проверить эффективность улучшений.

Порядок выполнения работы

Получив у преподавателя исходные данные для выполнения лабораторной работы, студент изучает методику построения диаграммы Парето. Далее строит диаграмму Парето на компьютере на основе исходных данных и делает анализ.

Содержание отчета

Отчет по лабораторной работе должен содержать следующие пункты:

- название лабораторной работы;
- цель работы;
- индивидуальное задание для выполнения лабораторной работы;
- краткие теоретические сведения;
- краткое описание хода выполнения работы;
- результаты выполнения работы: таблица с исходными данными, общая диаграмма Парето, диаграммы Парето для рабочего А и В, таблицы построения диаграмм Парето.
- выводы о качестве работы рабочих.

Контрольные вопросы

1. Что такое диаграмма Парето и для чего она применяется?
2. Назовите виды диаграмм Парето?
3. Какова область применения диаграммы Парето?
4. Каково значения диаграммы Парето в статистическом контроле качества продукции?

Лабораторная работа №4

Применение корреляционного и регрессионного анализа для установления причин дефектности продукции

Цель работы: получить практические навыки в обработке статистических данных, построении диаграммы разброса, проведении корреляционного и регрессионного анализа в программе *STATISTICA* или *LibreOffice* для установления причин дефектности продукции.

Задание для самостоятельного выполнения

Фирма-производитель пластиковых емкостей, изготавливаемых методом литья под давлением, сталкивается с трудностями из-за дефектных емкостей, имеющих слишком тонкие стенки. Было высказано предположение, что, причина неподходящей толщины стенок заключается в вариации давления сжатого воздуха, которое каждый день меняется. Известны данные о давлении воздуха и доли дефектов. По исходным данным полученным у преподавателя:

1. Построить диаграмму разброса и исследовать ее.
2. Провести корреляционный анализ и установить причины дефектности продукции. Проверить значимость корреляции.
3. Провести регрессионный анализ и установить причины дефектности продукции. Проверить значимость регрессии.

Краткие теоретические сведения

В процессе исследования часто приходится выяснять, существует ли зависимость между двумя различными параметрами процесса. Например, зависит ли качество готового изделия от качества исходных материалов, комплектующих деталей и узлов и т.д. Для выяснения зависимости между показателями качества и основными факторами производства, а также корреляционной зависимости между факторами используют диаграммы разброса (рассеивания), которые также называются полем корреляции.

Диаграмма разброса (рассеивания) – это инструмент, позволяющий определить вид и тесноту связи двух рассматриваемых параметров процесса.

Диаграмма разброса представляет собой график, получаемый путем нанесения в определенном масштабе экспериментальных, полученных в результате наблюдений точек. Координаты точек

соответствуют значениям рассматриваемой величины и влияющего на него фактора. Расположение точек на графике показывает наличие и характер связи между случайными величинами. Таким образом, диаграмма разброса дает возможность выдвинуть гипотезу о наличии или отсутствии корреляционной связи между двумя случайными величинами, которые могут относиться к характеристике качества и влияющему на нее фактору либо к двум различным характеристикам качества, либо к двум факторам, влияющим на одну характеристику качества.

По полученным экспериментальным точкам могут быть определены и числовые характеристики связи между рассматриваемыми случайными величинами: коэффициент корреляции и коэффициенты регрессии.

Корреляционный анализ

В связи с тем, что явления взаимосвязаны, одной из задач математической статистики является установление и измерение причинно-следственных связей между изучаемой случайной величиной y и одной или несколькими случайными величинами x_1, x_2, \dots, x_n . Изменение средней величины одного из признаков происходит в зависимости от изменения значения другого признака - статистическая зависимость подобного рода называется корреляционной зависимостью.

При наличии корреляционной зависимости можно осуществить контроль только одной (любой) из двух характеристик. При этом характер корреляционной зависимости, который определяется видом диаграммы разброса, дает представление о том, каким изменениям будет подвержен один из параметров при определенных изменениях другого. Так, при увеличении x на диаграмме (см. рисунок 4.1а) y также будет увеличиваться (прямая корреляция). В этом случае при осуществлении контроля причинных факторов x характеристика y (функция) будет оставаться стабильной.

На рисунке 4.1б показан пример обратной (отрицательной) корреляции. При увеличении x характеристика y уменьшается. Если причинный фактор x находится под контролем, характеристика y остается стабильной.

На рисунке 4.1в показан пример отсутствия корреляции, когда

никакой выраженной зависимости между x и y не наблюдается. В этом случае необходимо продолжить поиск факторов, коррелирующих с y , исключив из этого поиска фактор x .

Между параметрами x и y возможны также случаи криволинейной корреляции (см. рисунок 4.1г). Если при этом диаграмму разброса можно разделить на участки, имеющие прямолинейный характер, проводят такое разделение и исследуют каждый участок в отдельности.

В некоторых случаях вывод, полученный на основе визуального анализа диаграмм рассеяния, бывает достаточным для принятия решений о проведении нужных мероприятий. Но иногда желательно получить количественную оценку тесноты или силы связи между случайными величинами.

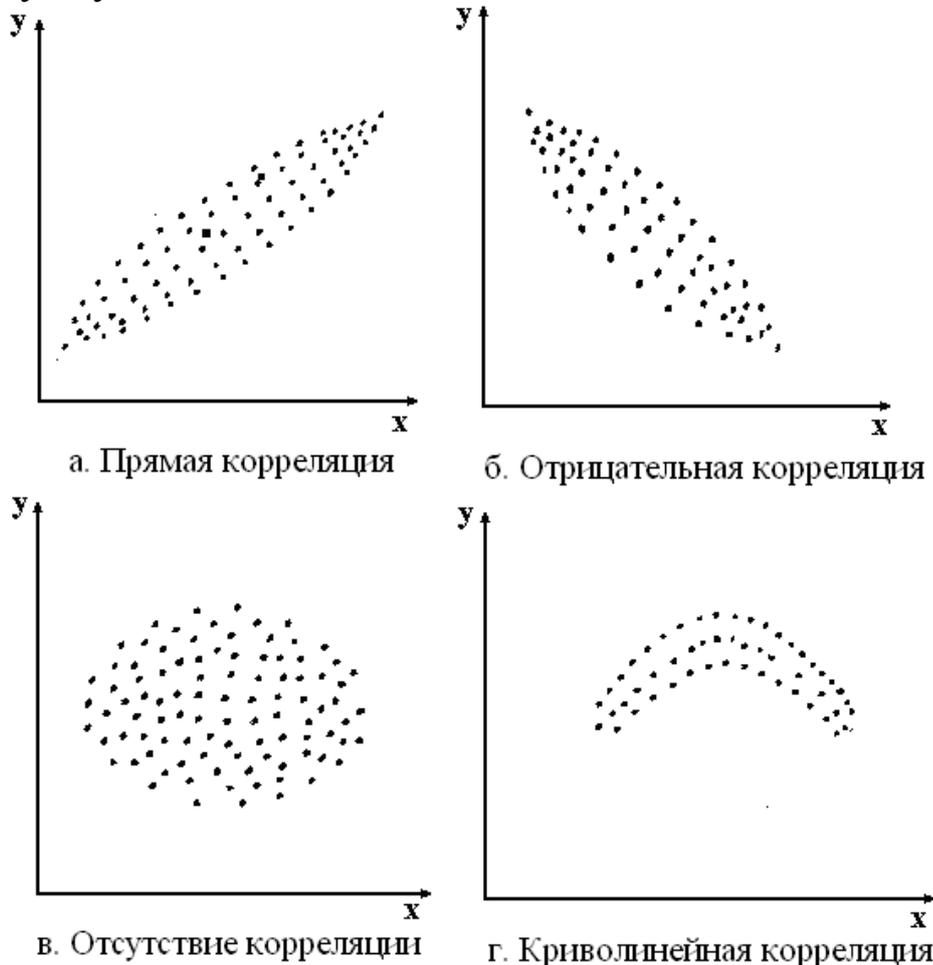


Рисунок 4.1 – Диаграммы разброса

Существуют различные методы оценки степени корреляционной зависимости. Одним из них является метод вычисления коэффициента корреляции r по формуле:

$$r = \frac{S(xy)}{\sqrt{S(xx)S(yy)}}, \quad (4.1)$$

где n – число пар данных

$$S(xx) = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \quad (4.2)$$

$$S(yy) = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} \quad (4.3)$$

$$S(xy) = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}, \quad (4.4)$$

а $S(xy)$ называется ковариацией.

Ковариация – это показатель тесноты связи между переменными.

Коэффициент корреляции r принимает значения в диапазоне $-1 \leq r \leq 1$. Если абсолютное значение r окажется больше 1, то совершенно ясно, что произошла ошибка, и необходимо пересчитать результат. Если коэффициент корреляции $|0 \div 0,3|$, это свидетельствует о слабой корреляции, а если $r = [0,3 \div 0,6|$ – средняя корреляция, $r = [0,6 \div 1,0|$ – сильная корреляция. Таким образом, в случае сильной положительной корреляции достигается значение, близкое к +1, а при сильной отрицательной корреляции достигается значение, близкое к -1. А, когда $|r|$ близок к 1, это указывает на сильную корреляцию между x и y , а когда $|r|$ близок к 0 – на слабую корреляцию. Более того, при $|r| = 1$ все данные будут лежать на прямой. Если $|r| = 0$ это свидетельствует об отсутствии статистической зависимости между переменными. При таком значении коэффициента корреляции регрессионный анализ не проводится.

Критерий значимости корреляции

Пусть r – выборочный коэффициент корреляции, вычисленный по выборке объема n из генеральной совокупности, имеющей нормальное распределение. Требуется на заданном

уровне значимости α проверить нулевую гипотезу о равенстве нулю коэффициента корреляции для генеральной совокупности $H_0: R = 0$.

Если нулевая гипотеза будет отвергнута, то говорят о значимости коэффициента корреляции, т.е. о наличии корреляции между X , Y . Если же нулевая гипотеза принимается, то корреляция не значима: X и Y некоррелированы (несмотря на то, что выборочный коэффициент корреляции $r \neq 0$).

Для проверки рассматриваемой гипотезы используется статистика

$$t = r \sqrt{\frac{n-2}{1-r^2}}, \quad (4.5)$$

имеющая распределение Стьюдента с числом степеней свободы $(n-2)$.

Регрессионный анализ. Линейная регрессия

Количественный подход к представлению связи между x и y , в виде линии регрессии, называется регрессионным анализом. Линии регрессии, представляет собой уравнение:

$$y = a_0 + a_1x \quad (4.6)$$

где y – зависимая переменная; x – независимая переменная (фактор); a_0 – константа (свободный член); a_1 – коэффициент регрессии (угловой коэффициент).

Величина коэффициента регрессии a_1 показывает, насколько в среднем изменится y при изменении x на единицу своего измерения. Знак коэффициента регрессии указывает на направление связи. Если $a_1 > 0$, то связь между переменными прямая, т.е. с увеличением x увеличивается y и наоборот. Если $a_1 < 0$, то связь между переменными обратная, т.е. с увеличением x параметр y уменьшается и наоборот.

Значение параметра a_0 в уравнении регрессии трактуется как среднее значение параметра y при условии, что $x=0$. Такая трактовка параметра a_0 возможна только в том случае, если значение $x=0$ имеет смысл.

Пусть (x_i, y_i) ($1 \leq i \leq n$) будет множеством из n пар результатов наблюдений, \hat{a}_0 и \hat{a}_1 – оценками значений a_0 и a_1 .

Этап 1. Найти по имеющимся данным \bar{y} и \bar{x} .

Этап 2. Подсчитайте $S(xx)$ и $S(xy)$.

Этап 3. Найдите значение \hat{a}_1 по формуле (4.7) и \hat{a}_0 по формуле (4.8):

$$\hat{a}_1 = \frac{S(xy)}{S(xx)}, \quad (4.7)$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x} \quad (4.8)$$

Этап 4. Подставляя полученные данные в уравнение, получаем уравнение линии регрессии.

Следует отметить, что прежде чем проводить корреляционный анализ, так же как и регрессионный, необходимо построить диаграмму разброса.

Критерий значимости регрессии и оценка ее адекватности

Задача регрессионного анализа помимо оценки параметров регрессии a_0 и a_1 сводится и к проверке гипотезы о значимости модели и оценке ее адекватности: достаточно ли хорошо согласуется модель (4.6) с результатами наблюдений?

Мерой значимости линии регрессии может служить следующее соотношение:

$$F[m, n - m - 1] = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 / (m \sigma_{YX}^2). \quad (4.9)$$

где m – число факторных признаков (независимых переменных).

Действительно, связь тем больше, чем значительнее мера рассеяния признака, обусловленная регрессией, превосходит меру рассеяния отклонений фактических значений от выравненных.

Тогда используя основное тождество дисперсионного анализа

$$S(yu) = S(R) + S(e) \quad (4.10)$$

где $S(yu)$ – общая сумма квадратов; $S(R)$ – сумма квадратов, обусловленная регрессией; $S(e)$ – остаточная сумма квадратов.

Соотношение (4.9) позволяет решить вопрос о значимости регрессии.

Например, парная линейная регрессионная модель называется незначимой, если параметр $a_1 = 0$. Для проверки нулевой гипотезы $H_0: a_1 = 0$ используется статистика (4.9) при $m = 1$:

$$F_\phi[1, n - 2] = \frac{S(R)}{\frac{1}{n - 2} S(e)}, \quad (4.11)$$

Регрессия значима, т. е. между признаками существует линейная связь, если для данного уровня значимости α вычисленное значение $F_{\text{ф}}[m, n-m-1]$ превышает критическое значение $F_{\text{кр}}[m, n-m-1]$, стоящее на пересечении m -го столбца и $[n-m-1]$ -й строки таблицы Фишера. Таким образом, нулевая гипотеза отклоняется: регрессионная модель статистически значима.

Кроме значимости, проверяется и адекватность модели. Приблизительно адекватность можно проверить по диаграмме разброса с нанесенной на нее расчетной прямой (см. рисунок 4.2).

Коэффициент детерминации

С помощью F-критерия можно установить, что существует линейная зависимость между величинами X и Y . Следовательно, можно утверждать, что величина Y линейно зависит от величины X . Теперь уместно поставить уточняющий вопрос - в какой степени величина X определяет величину Y ? На этот вопрос можно ответить, рассчитав, какая часть вариации результативного признака Y может быть объяснена влиянием факторного признака X .

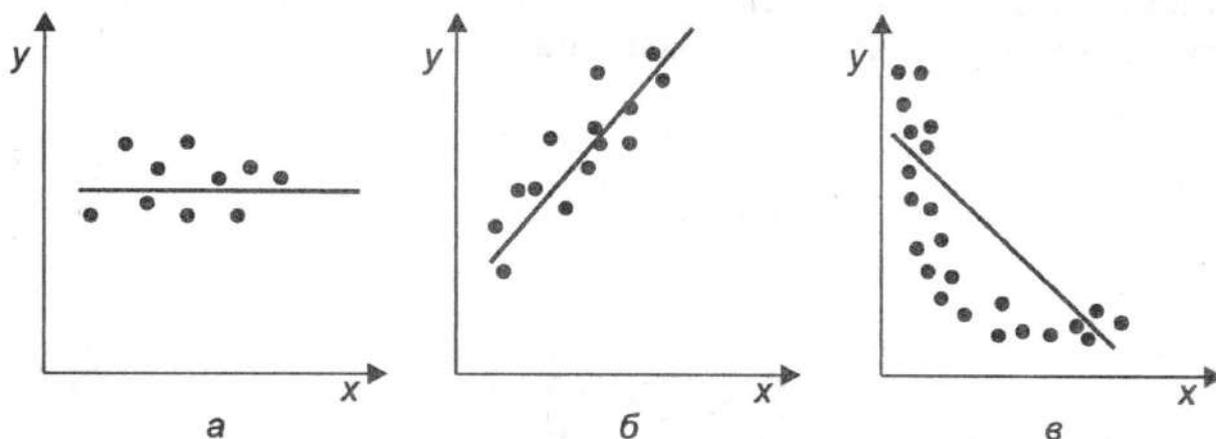


Рисунок 4.2 – Значимость и адекватность парной линейной регрессии: а) модель незначима; б) модель значима и адекватна; в) модель значима, но неадекватна

Рассмотрим отношение

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}. \quad (4.12)$$

Оно показывает долю разброса, учитываемого регрессией, в

общем разбросе результативного признака и носит название коэффициента детерминации. Этот показатель, равный отношению факторной вариации к полной вариации признака, позволяет судить о том, насколько «удачно» выбран вид функции. Проведя расчеты, основанные на одних и тех же исходных данных, для нескольких типов функций, мы можем выбрать такую, которая дает наибольшее значение R^2 и, следовательно, в большей степени, чем другие функции, объясняет вариацию результативного признака. Действительно, при расчете R^2 для одних и тех же данных, но разных функций знаменатель выражения (12) остается неизменным, а числитель показывает ту часть вариации результативного признака, которая учитывается выбранной функцией. Чем больше R^2 , т. е. чем больше числитель, тем больше изменение факторного признака объясняет изменение результативного признака и тем, следовательно, лучше уравнение регрессии, лучше выбор функции.

Случай двух независимых переменных. Простейший случай множественной регрессии

В предыдущем изложении регрессионного анализа мы имели дело с двумя признаками — результативным и факторным. Но на результат действует обычно не один фактор, а несколько, что необходимо учитывать для достаточно полного анализа связей.

В математической статистике разработаны методы множественной регрессии (Регрессия называется множественной, если число независимых переменных, учтенных в ней, больше или равно двум.), позволяющие анализировать влияние на результативный признак нескольких факторных.

Простейшее уравнение множественной регрессии будет выглядеть так:

$$Y = a_0 + a_1X_1 + a_2X_2, \quad (4.13)$$

где Y — результативный признак или зависимая переменная; X_1 и X_2 — независимые переменные; a_0 , a_1 , a_2 — параметры, подлежащие определению.

Для нахождения числовых значений искомых параметров, как и в случае одной независимой переменной, пользуются методом наименьших квадратов. Он сводится к составлению и решению системы нормальных уравнений, которая имеет вид:

$$\begin{cases} na_0 + a_1 \sum_{i=1}^n X_{1i} + a_2 \sum_{i=1}^n X_{2i} = \sum_{i=1}^n Y_i \\ a_0 \sum_{i=1}^n X_{1i} + a_1 \sum_{i=1}^n X_{1i}^2 + a_2 \sum_{i=1}^n X_{1i} X_{2i} = \sum_{i=1}^n Y_i X_{1i} \\ a_0 \sum_{i=1}^n X_{2i} + a_1 \sum_{i=1}^n X_{1i} X_{2i} + a_2 \sum_{i=1}^n X_{2i}^2 = \sum_{i=1}^n Y_i X_{2i} \end{cases} .$$

Когда система состоит из трех и более нормальных уравнений, решение ее усложняется. Существуют стандартные программы расчета неизвестных параметров регрессионного уравнения на ЭВМ. При ручном счете можно воспользоваться методом Гаусса.

Интерпретация коэффициентов уравнения множественной регрессии

Коэффициент при независимой переменной в уравнении простой регрессии отличается от коэффициента при соответствующей переменной в уравнении множественной регрессии тем, что в последнем элиминировано влияние всех учтенных в данном уравнении признаков.

Коэффициенты уравнения множественной регрессии, поэтому называются частными или чистыми коэффициентами регрессии.

Величина коэффициентов регрессии изменяется в зависимости от единиц измерения, в которых представлены переменные. Если переменные выражены в разном масштабе измерения, то соответствующие им коэффициенты становятся несравнимыми. Для достижения сопоставимости коэффициенты регрессии исходного уравнения стандартизуют, взяв вместо исходных переменных их отношения к собственным средним квадратическим отклонениям. Тогда уравнение (4.13) приобретает вид

$$\frac{Y}{\sigma_Y} = a'_0 + \beta_1 \frac{X_1}{\sigma_{X_1}} + \beta_2 \frac{X_2}{\sigma_{X_2}} \quad (4.14)$$

Сравнивая полученное уравнение с уравнением (4.13), определяем стандартизованные частные коэффициенты уравнения, так называемые бета-коэффициенты, по формулам:

$$\beta_1 = a_1 \frac{\sigma_{X_1}}{X_1}, \quad \beta_2 = a_2 \frac{\sigma_{X_2}}{X_2}, \quad (4.15)$$

где β_1 и β_2 — бета-коэффициенты; a_1 и a_2 — коэффициенты регрессии исходного уравнения; σ_Y , σ_{X_1} и σ_{X_2} — средние квадратические отклонения переменных Y , X_1 и X_2 соответственно.

Оценка полезности введения дополнительной переменной

Точность уравнения регрессии тесно связана с вопросом ценности включения дополнительных членов в это уравнение.

Включение в уравнение новой переменной приводит к изменению средней квадратической ошибки. Если средняя квадратическая ошибка уменьшается, то это уже может говорить о полезности введения дополнительной переменной.

Также по формуле (4.12) можно рассчитать коэффициент детерминации R^2 .

Если при включении в уравнение новой переменной происходит увеличение R^2 , то это говорит об увеличении точности уравнения.

Таким образом, сравнение средних квадратических ошибок уравнения, коэффициентов детерминации, рассчитанных до и после введения независимой переменной, позволяет судить о полезности включения этой переменной в уравнение. Однако следует быть осторожными в выводах при подобных сравнениях, поскольку увеличение R^2 или уменьшение σ не всегда имеют приписываемый им здесь смысл. Так, увеличение R^2 может объясняться тем фактом, что число рассматриваемых параметров в уравнении приближается к числу объектов наблюдения. Скажем, весьма сомнительными будут ссылки на увеличение R^2 или уменьшение σ , если в уравнение вводится третья или четвертая независимая переменная и уравнение строится на данных по шести, семи объектам.

Полезность включения дополнительного фактора можно оценить с помощью F-критерия (4.9).

Частный F-критерий показывает степень влияния дополнительной независимой переменной на результативный признак и может использоваться при решении вопроса о добавлении в уравнение или исключении из него этой независимой переменной.

Разброс признака, объясняемый уравнением регрессии (4.13), можно разложить на два вида: 1) разброс признака, обусловленный

независимой переменной X_1 , и 2) разброс признака, обусловленный независимой переменной X_2 , когда X_1 уже включена в уравнение. Первой составляющей соответствует разброс признака, объясняемый уравнением (4.6), включающим только переменную X_1 . Разность между разбросом признака, обусловленным уравнением (4.13), и разбросом признака, обусловленным уравнением (4.6), определит ту часть разброса, которая объясняется дополнительной независимой переменной X_2 . Отношение указанной разности к разбросу признака, регрессией не объясняемому, представляет собой значение частного критерия. Частный F-критерий называется также последовательным, если статистические характеристики строятся при последовательном добавлении переменных в регрессионное уравнение.

Важным условием применения к обработке данных метода множественной регрессии является отсутствие сколько-нибудь значительной взаимосвязи между факторными признаками. При практическом использовании метода множественной регрессии, прежде чем включать факторы в уравнение, необходимо убедиться в том, что они независимы.

Если один из факторов зависит линейно от другого, то система нормальных уравнений, используемая для нахождения параметров уравнения, не разрешима. Содержательно этот факт можно толковать так: если факторы X_1 и X_2 связаны между собой, то они действуют на результативный признак у практически как один фактор, т. е. сливаются воедино и их влияние на изменение Y разделить невозможно. Когда между независимыми переменными уравнения множественной регрессии имеется линейная связь, следствием которой является неразрешимость системы нормальных уравнений, то говорят о наличии мультиколлинеарности.

На практике вопрос о наличии или об отсутствии мультиколлинеарности решается с помощью показателей взаимосвязи. В случае двух факторных признаков используется парный коэффициент корреляции между ними: если этот коэффициент по абсолютной величине превышает 0.8, то признаки относят к числу мультиколлинеарных. Если число факторных признаков больше двух, то рассчитываются множественные коэффициенты корреляции. Фактор признается

мультиколлинеарным, если множественный коэффициент корреляции, характеризующий совместное влияние на этот фактор остальных факторных признаков, превзойдет по величине коэффициент множественной корреляции между результативным признаком и совокупностью всех независимых переменных.

Самый естественный способ устранения мультиколлинеарности — исключение одного из двух линейно связанных факторных признаков. Этот способ прост, но не всегда приемлем, так как подлежащий исключению фактор может оказывать на зависимую переменную особое влияние. В такой ситуации применяются более сложные методы избавления от мультиколлинеарности.

Выбор «наилучшего» уравнения регрессии

Эта проблема связана с двойственным отношением к вопросу о включении в регрессионное уравнение независимых переменных. С одной стороны, естественно стремление учесть все возможные влияния на результативный признак и, следовательно, включить в модель полный набор выявленных переменных. С другой стороны, возрастает сложность расчетов и затраты, связанные с получением максимума информации, могут оказаться неоправданными. Нельзя забывать и о том, что для построения уравнения регрессии число объектов должно в несколько раз превышать число независимых переменных. Эти противоречивые требования приводят к необходимости компромисса, результатом которого и является «наилучшее» уравнение регрессии. Существует несколько методов, приводящих к цели: метод всех возможных регрессий, метод исключения, метод включения, шаговый регрессионный и ступенчатый регрессионный методы.

Метод всех возможных регрессий заключается в переборе и сравнении всех потенциально возможных уравнений. В качестве критерия сравнения используется коэффициент детерминации R^2 . «Наилучшим» признается уравнение с наибольшей величиной R^2 . Метод весьма трудоемок и предполагает использование вычислительных машин.

Методы исключения и включения являются усовершенствованными вариантами предыдущего метода. В методе исключения в качестве исходного рассматривается регрессионное

уравнение, включающее все возможные переменные. Рассчитывается частный F-критерий для каждой из переменных, как будто бы она была последней переменной, введенной в регрессионное уравнение. Минимальная величина частного F-критерия (F_{\min}) сравнивается с критической величиной ($F_{кр}$), основанной на заданном исследователем уровне значимости. Если $F_{\min} > F_{кр}$, то уравнение остается без изменения. Если $F_{\min} \leq F_{кр}$, то переменная, для которой рассчитывался этот частный F-критерий, исключается. Производится перерасчет уравнения регрессии для оставшихся переменных, и процедура повторяется для нового уравнения регрессии. Исключение из рассмотрения уравнений с незначимыми переменными уменьшает объем вычислений, что является достоинством этого метода по сравнению с предыдущим.

Метод включения состоит в том, что в уравнение включаются переменные по степени их важности до тех пор, пока уравнение не станет достаточно «хорошим». Степень важности определяется линейным коэффициентом корреляции, показывающим тесноту связи между анализируемой независимой переменной и результативным признаком: чем теснее связь, тем больше информации о результирующем признаке содержит данный факторный признак и тем важнее, следовательно, введение этого признака в уравнение.

Процедура начинается с отбора факторного признака, наиболее тесно связанного с результативным признаком, т. е. такого факторного признака, которому соответствует максимальный по величине парный линейный коэффициент корреляции. Далее строится линейное уравнение регрессии, содержащее отобранную независимую переменную. Выбор следующих переменных осуществляется с помощью частных коэффициентов корреляции, в которых исключается влияние вошедших в модель факторов. Для каждой введенной переменной рассчитывается частный F-критерий, по величине которого судят о том, значим ли вклад этой переменной. Как только величина частного F-критерия, относящаяся к очередной переменной, оказывается незначимой, т. е. эффект от введения этой переменной становится малозаметным, процесс включения переменных заканчивается. Метод включения связан с меньшим объемом

вычислений, чем предыдущие методы. Но при введении новой переменной нередко значимость включенных ранее переменных изменяется. Метод включения этого не учитывает, что является его недостатком. Модификацией метода включения, исправляющей этот недостаток, является шаговый регрессионный метод.

Шаговый регрессионный метод кроме процедуры метода включения содержит анализ переменных, включенных в уравнение на предыдущей стадии. Потребность в таком анализе возникает в связи с тем, что переменная, обоснованно введенная в уравнение на ранней стадии, может оказаться лишней из-за взаимосвязи ее с переменными, позднее включенными в уравнение. Анализ заключается в расчете на каждом этапе частных F -критериев для каждой переменной уравнения и сравнении их с величиной $F_{кр}$, точкой F -распределения, соответствующей заданному исследователем уровню значимости. Частный F -критерий показывает вклад переменной в вариацию результативного признака в предположении, что она вошла в модель последней, а сравнение его с $F_{кр}$ позволяет судить о значимости рассматриваемой переменной с учетом влияния позднее включенных факторов. Незначимые переменные из уравнения исключаются.

Рассмотренные методы предполагают довольно большой объем вычислений и практически неосуществимы без ЭВМ. Для реализации ступенчатого регрессионного метода вполне достаточно малой вычислительной техники.

Ступенчатый регрессионный метод включает в себя такую последовательность действий. Сначала выбирается наиболее тесно связанная с результативным признаком переменная и составляется уравнение регрессии. Затем находят разности фактических и выравненных значений и эти разности (остатки) рассматриваются как значения результативной переменной. Для остатков подбирается одна из оставшихся независимых переменных и т. д. На каждой стадии проверяется значимость регрессии. Как только обнаружится незначимость, процесс прекращается и окончательное уравнение получается суммированием уравнений, полученных на каждой стадии за исключением последней.

Ступенчатый регрессионный метод менее точен, чем

предыдущие, но не столь громоздок. Он оказывается полезным в случаях, когда необходимо внести содержательные правки в уравнение. Так, для изучения факторов, влияющих на цены угля в Санкт-Петербурге в конце XIX— начале XX в., было получено уравнение множественной регрессии. В него вошли следующие переменные: цены угля в Лондоне, добыча угля в России и экспорт из России. Здесь не обосновано появление в модели такого фактора, как добыча угля, поскольку Санкт-Петербург работал исключительно на импортном угле. Модели легко придать экономический смысл, если независимую переменную «добыча» заменить независимой переменной «импорт». Формально такая замена возможна, поскольку между импортом и добычей существует тесная связь. Пользуясь ступенчатым методом, исследователь может совершить эту замену, если предпочтет содержательно интерпретируемый фактор.

Нелинейная регрессия

Построение уравнений нелинейной регрессии

До сих пор мы, в основном, изучали связи, предполагая их линейность. Но не всегда связь между признаками может быть достаточно хорошо представлена линейной функцией. Иногда для описания существующей связи более пригодными, а порой и единственно возможными являются более сложные нелинейные функции. Ограничимся рассмотрением наиболее простых из них.

Одним из простейших видов нелинейной зависимости является парабола, которая в общем виде может быть представлена функцией:

$$Y = a_0 + a_1 X + a_2 X^2. \quad (4.16)$$

Неизвестные параметры a_0 , a_1 , a_2 находятся в результате решения следующей системы уравнений:

$$\begin{cases} na_0 + a_1 \sum_{i=1}^n X_i + a_2 \sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i, \\ a_0 \sum_{i=1}^n X_i + a_1 \sum_{i=1}^n X_i^2 + a_2 \sum_{i=1}^n X_i^3 = \sum_{i=1}^n Y_i X_i, \\ a_0 \sum_{i=1}^n X_i^2 + a_1 \sum_{i=1}^n X_i^3 + a_2 \sum_{i=1}^n X_i^4 = \sum_{i=1}^n Y_i X_i^2. \end{cases} \quad (4.17)$$

Дает ли преимущества описание связи с помощью параболы по сравнению с описанием, построенным по гипотезе линейности? Ответ на этот вопрос можно получить, рассчитав последовательный F-критерий.

На практике для изучения связей используются полиномы более высоких порядков (3-го и 4-го порядков). Составление системы, ее решение, а также решение вопроса о полезности повышения порядка функции для этих случаев аналогичны описанным. При этом никаких принципиально новых моментов не возникает, но существенно увеличивается объем расчетов.

Кроме класса парабол для анализа нелинейных связей можно применять и другие виды функций. Для расчета неизвестных параметров этих функций рекомендуется использовать метод наименьших квадратов, как наиболее мощный и широко применяемый.

Однако метод наименьших квадратов не универсален, поскольку он может использоваться только при условии, что выбранные для выравнивания функции линейны по отношению к своим параметрам. Не все функции удовлетворяют этому условию, но большинство применяемых на практике с помощью специальных преобразований могут быть приведены к стандартной форме функции с линейными параметрами.

Измерение тесноты связи при криволинейной зависимости

Рассмотренные ранее линейные коэффициенты корреляции оценивают тесноту взаимосвязи при линейной связи между признаками. При наличии криволинейной связи указанные меры связи не всегда приемлемы.

Для измерения тесноты связи при криволинейной зависимости используется индекс корреляции, вычисляемый по формуле

$$i = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}}, \quad (4.18)$$

где Y_i – i -е значение результативного признака; \hat{Y}_i – i -е выравненное значение этого признака; \bar{Y} – среднее арифметическое значение результативного признака.

Числитель формулы (4.18) характеризует разброс выравненных значений результативного признака. Поскольку изменения выравненных, т. е. вычисленных по уравнению регрессии, значений признака происходят только в результате изменения факторного признака X , то числитель измеряет разброс результативного признака, обусловленный влиянием на него факторного признака. Знаменатель же измеряет разброс признака-результата, который определен влиянием на него всех факторов, в том числе и учтенного. Таким образом, индекс корреляции оценивает участие данного факторного признака в общем действии всего комплекса факторов, вызывающих колеблемость результативного признака, тем самым определяя тесноту зависимости признака Y от признака X . При этом, если признак X не вызывает никаких изменений признака Y , то числитель и, следовательно, индекс корреляции равны 0. Если же линия регрессии полностью совпадает с фактическими данными, т. е. признаки связаны функционально, то индекс корреляции равен 1. В случае линейной зависимости между X и Y индекс корреляции численно равен линейному коэффициенту корреляции r . Квадрат индекса корреляции совпадает с введенным ранее (4.12) коэффициентом детерминации.

Таковы основные принципы и условия, методика и техника применения регрессионного анализа.

Выполнение корреляционного и регрессионного анализа в программе STATISTICA

Исследование параметров корреляционной и регрессионной зависимости случайных величин проиллюстрируем на **примере**: имеются результаты наблюдения двух случайных x и y величин для них требуется определить характеристики распределения, построить график рассеивания, корреляционную матрицу, параметры уравнения регрессии.

Запускаем программу STATISTICA.

Создаем новую таблицу.

Задаём параметры новой таблицы – число строк 30 и столбцов 2. Для первого столбца задаём параметры следующим образом. В окне описания переменных задаём имя « X », в этом же окне задаём название второму столбцу « Y » (рисунок 4.3).

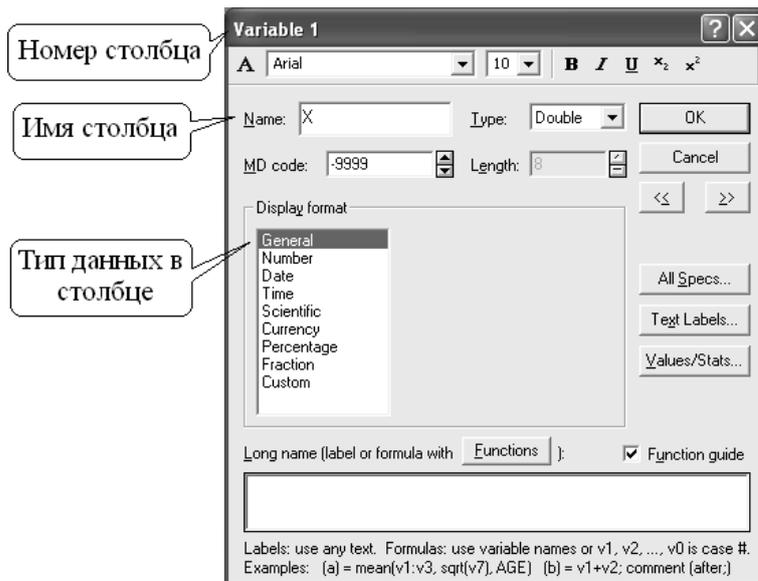


Рисунок 4.3 – Окно описания переменных

Заполняем столбцы переменных данными наблюдений.

Открываем окно исследования корреляционных зависимостей. В программе выделяем следующие вкладки **Статистика** → **Основная статистика**. В появившемся окне нажимаем на строку «Correlation matrices» (рисунок 4.4).

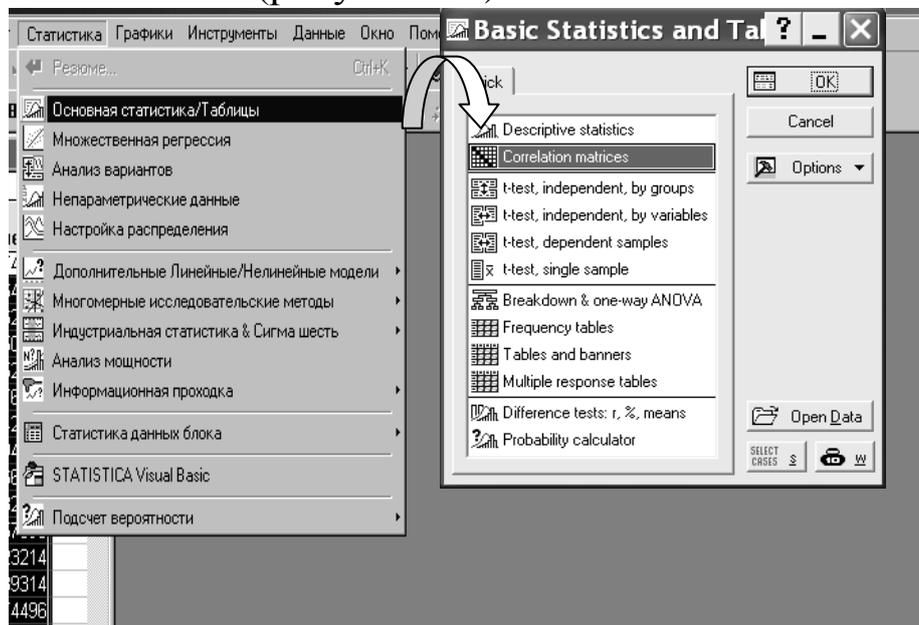


Рисунок 4 – Вкладка базовой статистики

В появившемся окне «Product-Moment and Partial Correlations» (рисунок 4.5) нажимаем на кнопку «One variable list». В следующем окне задаём переменные для вычислений (X, Y), удерживая клавишу Ctrl.

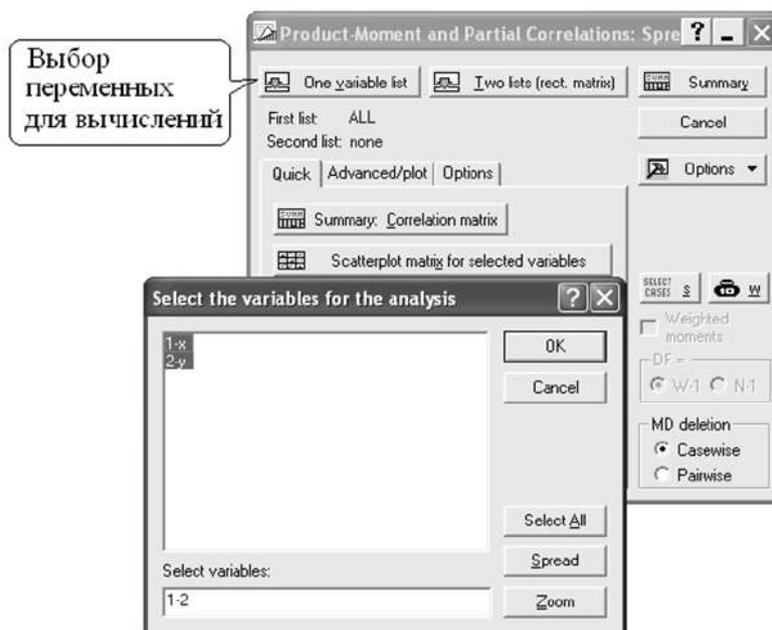


Рисунок 4.5 – Задание переменных для вычислений
Построение графиков рассеивания

Для этого в окне «Product-Moment and Partial Correlations» (рисунок 4.5) нажимаем на кнопку «Scatterplot matrix for selected variables» (рисунок 4.6).

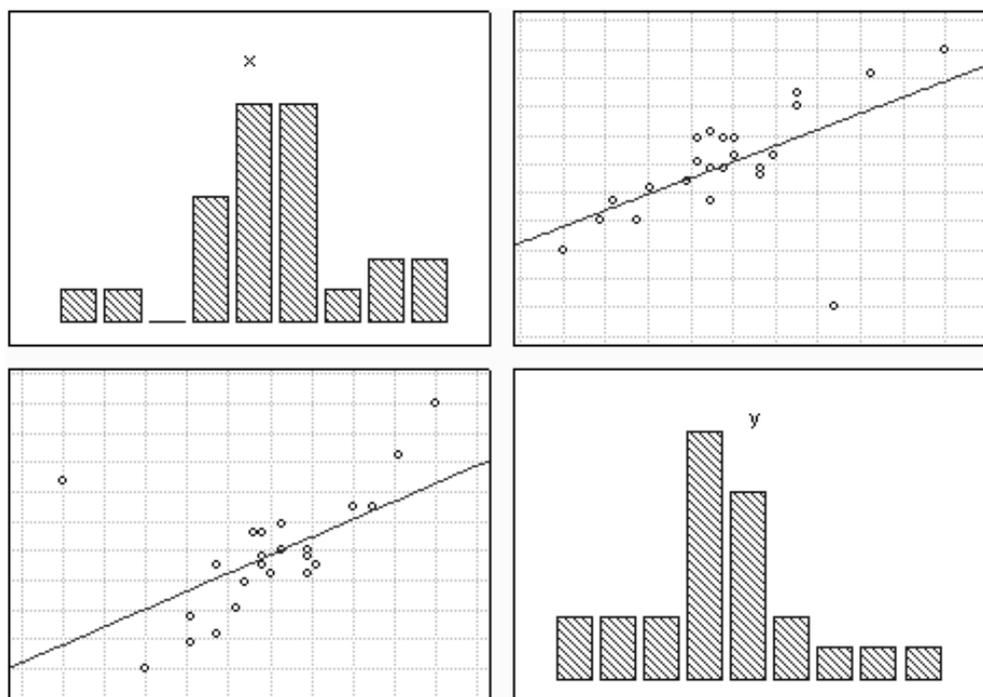


Рисунок 4.6 – Матрица графиков рассеивания

На основании графиков можно предположить о наличии положительной корреляционной зависимости между переменными.

Построение корреляционной матрицы

Для этого в окне «Product-Moment and Partial Correlations» (рисунок 4.5) нажимаем на кнопку « Summary: Correlation matrix» и появляется окно корреляционной матрицы (рисунок 4.7).

Correlations (Spreadsheet2)			
Marked correlations are significant at $p < ,05000$			
N=25 (Casewise deletion of missing data)			
Variable	x	y	
x	1,00	0,58	
y	0,58	1,00	

Кoeffициент корреляции

Рисунок 4.7 – Корреляционная матрица

Корреляционная матрица подтверждает, что между переменными существует положительная корреляционная связь ($r = 0,58$)

Проводим регрессионный анализ линейной модели

Выполняем следующие действия: **Статистика** →

Дополнительные Линейные/ Нелинейные модели → **Основные модели регрессии.**

При этом загрузится стартовое окно «General Regression Models» построения основных регрессионных моделей (рисунок 4.4).

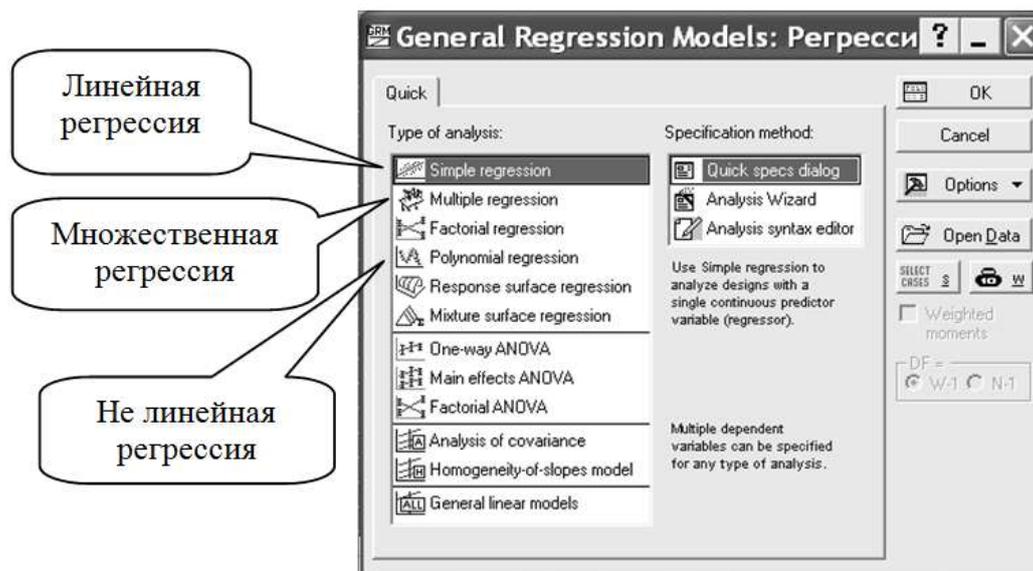


Рисунок 4.4 – Окно построения основных регрессионных моделей

Построение линейной модели

В окне «General Regression Models (Startup Panel)» (рисунок 4.4) на панели «Type of analysis» выбираем пункт «Simple

Regression», а на панели «Specification method» выбираем «Quick specs dialog» и нажимаем кнопку «ОК». В открывшемся окне «GRM Simple Regression» на вкладке «Quick» нажимаем кнопку «Variables» и указываем зависимую и не зависимую переменную (рисунок 4.5)

В окне «GRM Simple Regression» нажимаем кнопку «ОК», при этом откроется окно результатов расчета «GRM Results» (рисунок 4.6).

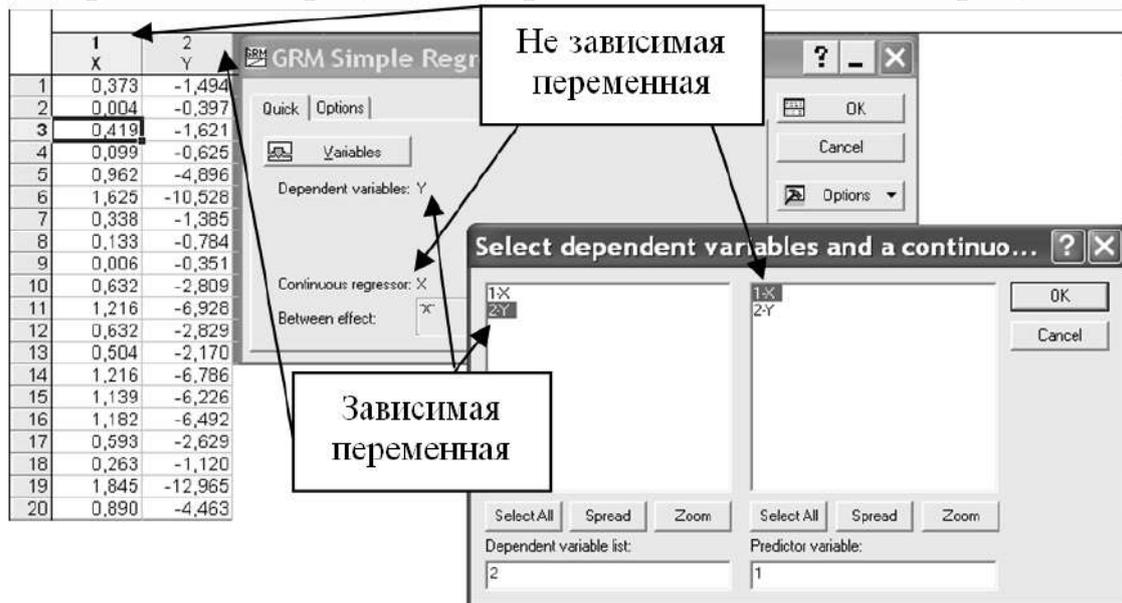


Рисунок 4.5 – Указание переменных для анализа



Рисунок 4.6 – Окно представления результатов анализа

Просматриваем результаты расчета, нажимаем кнопку «All effects». При этом будет построено три таблицы: «Univariate Tests of Significance for Y (Регрессия1)» (рисунок 4.7); «Parameter Estimates (Регрессия1)» (рисунок 4.8) и «Test of SS Whole Model vs. SS Residual (Регрессия1)» (рисунок 4.9). Уравнение линии регрессии имеет вид

$$Y=0,64953-6,43011*X.$$

Модель имеет не плохие показатели по коэффициентам, все вероятности ниже уровня значимости 0,1. Для визуального анализа точности модели построим диаграмму рассеивания на которой сопоставим опытную величину Y с величиной вычисленной с помощью уравнения регрессии \hat{Y} . Чем точнее

модель, тем ближе по значению Y и \hat{Y} , а значит точки на построенной диаграмме ближе расположены вдоль одной линии. Для построения данной диаграммы в программе STATSTICA в окне «GRM Results» (рисунок 4.6) открываем вкладку «Resids» (остатки) и нажимаем кнопку «Obs. & pred».

Effect	SS	Degr. of Freedom	MS	F	p
Intercept	2,9941	1	2,9941	5,8912	0,025936
X	225,1739	1	225,1739	443,0610	0,000000
Error	9,1480	18	0,5082		

Рисунок 4.7 – Таблица «Univariate Tests of Significance for Y»

Effect	Y Param.	Y Std. Err.	Y t	Y p	-95,00% Cnf. Lmt	+95,00% Cnf. Lmt	Y Beta (?)	Y St. Err. ?	-95,00% Cnf. Lmt	+95,00% Cnf. Lmt
Intercept	0,64953	0,267604	2,4272	0,025936	0,08731	1,21174				
X	-6,43011	0,305483	-21,0490	0,000000	-7,07191	-5,78832	-0,980285	0,046572	-1,07813	-0,882442

Рисунок 4.8 – Таблица «Parameter Estimates (Регрессия1)»

Dependent Variable	Multiple R	Multiple R?	Adjusted R?	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
Y	0,980285	0,960960	0,958791	225,1739	1	225,1739	9,148021	18	0,508223	443,0610	0,000000

Рисунок 4.9 – Таблица «Test of SS Whole Model vs. SS Residual (Регрессия1)»

На диаграмме (рисунок 4.10) отчетливо видно, что точки сильно отклоняются от прямой.

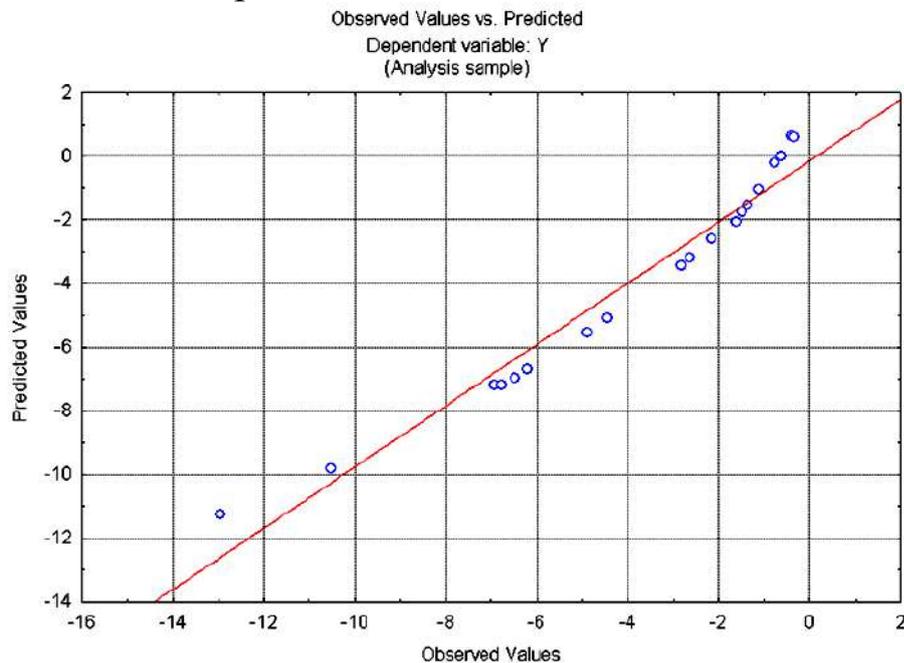


Рисунок 4.10 – Диаграмма рассеивания опытной величины Y с величиной высчитанной с помощью уравнения регрессии \hat{Y}

Построение нелинейной модели

Далее построим нелинейного уравнения регрессии (4.16), которое является полиномом 2-го порядка. Рассчитаем его характеристики, сопоставим их с характеристиками линейной модели и сделаем вывод о целесообразности перехода к нелинейной модели.

Затем последовательно будем повышать порядок полинома в уравнении регрессии и рассчитывать его характеристики до тех пор пока новое уравнение и его характеристики не окажется хуже предыдущего. Тогда предыдущее уравнение и будет являться решением поставленной задачи.

Построение уравнения $Y = a_0 + a_1X + a_2X^2$

Закрываем окна «GRM Results», а в окне «General Regression Models» (см. рисунок 4.4) на панели «Type of analysis» выбираем пункт «Polynomial Regression» и нажимаем кнопку «ОК». Задаем переменные для анализа в окне «GRM Polynomial Regression» точно также как при построении линейной модели. Получаем результаты расчета, строим три таблицы и диаграмму рассеивания (рисунок 4.11).

Univariate Tests of Significance for Y (Регрессия1) Sigma-restricted parameterization Effective hypothesis decomposition					
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	0,468290	1	0,468290	119,895	0,000000
X	2,909640	1	2,909640	744,948	0,000000
X^2	9,081622	1	9,081622	2325,144	0,000000
Error	0,066399	17	0,003906		

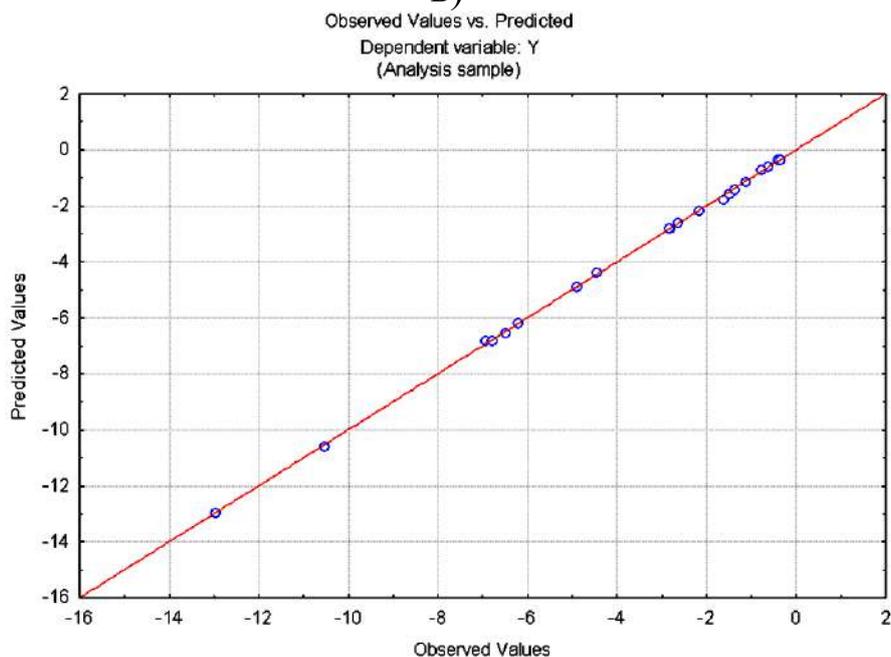
a)

Parameter Estimates (Регрессия1) Sigma-restricted parameterization										
Effect	Y Param.	Y Std.Err	Y t	Y p	-95,00% Cnf.Lmt	+95,00% Cnf.Lmt	Y Beta (?)	Y St.Err.?	-95,00% Cnf.Lmt	+95,00% Cnf.Lmt
Intercept	-0,34151	0,031189	-10,9497	0,000000	-0,40731	-0,27571				
X	-2,39701	0,087823	-27,2937	0,000000	-2,58230	-2,21172	-0,365430	0,013389	-0,393678	-0,337182
X^2	-2,40653	0,049908	-48,2197	0,000000	-2,51183	-2,30124	-0,645604	0,013389	-0,673852	-0,617356

б)

Test of SS Whole Model vs. SS Residual (Регрессия1)											
Dependent Variable	Multiple R	Multiple R?	Adjusted R?	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
Y	0,999858	0,999717	0,999683	234,2556	2	117,1278	0,066399	17	0,003906	29987,92	0,00

в)



г)

Рисунок 4.11 – Результаты построения уравнения

$$Y = a_0 + a_1X + a_2X^2$$

На основе полученных данных можно сделать следующие выводы:

– уравнение регрессии имеем вид $Y = -0,34151 - 2,39701 \cdot X - 2,40653 \cdot X^2$ (рисунок 4.11б), при этом параметры уравнения определены с достаточно малой величиной ошибки;

– значения частного F-критерия для параметров уравнения, по сравнению с линейной моделью (рисунок 4.7), увеличились (рисунок 4.11а), значит введение дополнительного члена в уравнение (X^2) было оправдано;

– коэффициенты корреляции и детерминации, а также мера значимости линии регрессии увеличились (рисунок 4.11в), по сравнению с линейной моделью (рисунок 4.9);

– на диаграмме рассеивания можно заметить, что точки стали расположены ближе к прямой линии (рисунок 4.11г), по сравнению с линейной моделью (рисунок 4.10).

Все выше перечисленное говорит о том, что уравнение регрессии $Y = a_0 + a_1X + a_2X^2$ лучше описывает взаимосвязь X с Y чем уравнение $Y = a_0 + a_1X$.

В окне «GRM Polynomial Regression» (рисунок 4.12) на вкладке «Quick» нажимаем кнопку «Between effects», при этом откроется окно «GLM Between Effects» в котором мы устанавливаем опцию «Use polynomials to specified degree» и величину порядка полинома «Degree: 3». Нажимаем кнопки «ОК» и как в предыдущих примерах получаем результаты расчета, строим три таблицы и диаграмму разброса (рисунок 4.13).

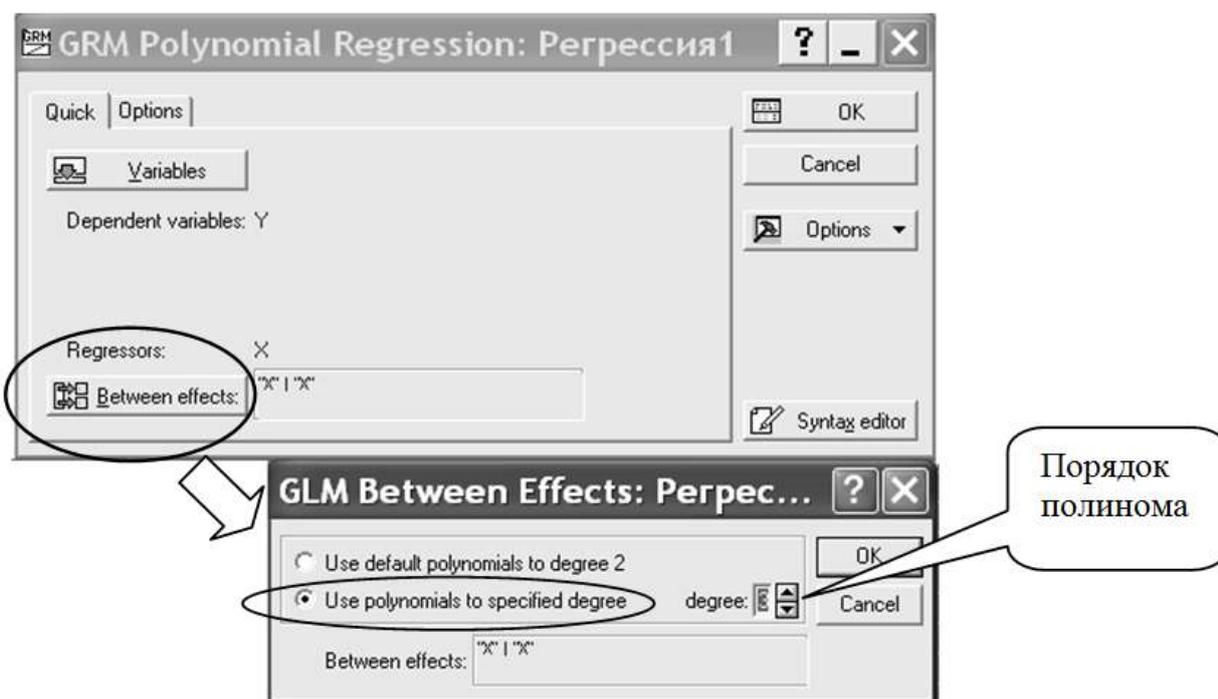


Рисунок 4.12 – Повышение порядка полинома

Univariate Tests of Significance for Y (Перессия1)
Sigma-restricted parameterization
Effective hypothesis decomposition

Effect	SS	Degr. of Freedom	MS	F	p
Intercept	0,386308	1	0,386308	105,3173	0,000000
X	0,476168	1	0,476168	129,8155	0,000000
X^2	0,432565	1	0,432565	117,9282	0,000000
X^3	0,007711	1	0,007711	2,1021	0,166416
Error	0,058689	16	0,003668		

a)

Parameter Estimates (Перессия1)
Sigma-restricted parameterization

Effect	Y	Y	Y	Y	-95,00%	+95,00%	Y	Y	-95,00%	+95,00%
	Param.	Std.Err	t	p	Cnf.Lmt	Cnf.Lmt	Beta (?)	St.Err.?	Cnf.Lmt	Cnf.Lmt
Intercept	-0,37001	0,036055	-10,2624	0,000000	-0,44644	-0,29358				
X	-2,15247	0,188919	-11,3937	0,000000	-2,55296	-1,75199	-0,328150	0,028801	-0,389205	-0,267094
X^2	-2,76960	0,255040	-10,8595	0,000000	-3,31026	-2,22894	-0,743003	0,068420	-0,888047	-0,597960
X^3	0,13555	0,093492	1,4499	0,166416	0,06264	0,33374	0,063191	0,043584	-0,029203	0,155585

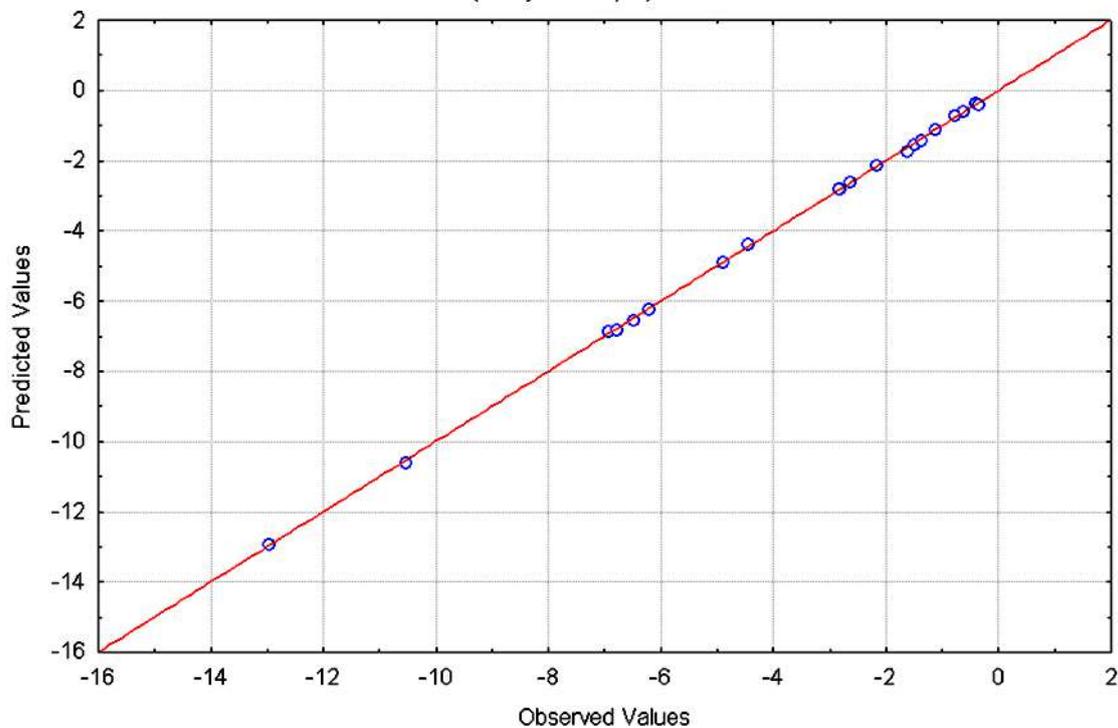
б)

Test of SS Whole Model vs. SS Residual (Перессия1)

Dependent Variable	Multiple R	Multiple R ²	Adjusted R ²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
Y	0,999875	0,999750	0,999703	234,2633	3	78,08776	0,058689	16	0,003668	21288,71	0,00

в)

Observed Values vs. Predicted

Dependent variable: Y
(Analysis sample)

г)

Рисунок 4.13 – Результаты построения уравнения

$$Y = a_0 + a_1X + a_2X^2 + a_3X^3$$

На основе полученных данных можно сделать следующие выводы:

- уравнение регрессии имеем вид $Y = -0,37001 - 2,15247 * X - 2,76960 * X^2 + 0,13555 * X^3$ (рисунок 4.13б), при этом величина ошибки определения частного коэффициента уравнения стоящего перед X^3 превышает критическое значение (вероятность $p = 0,166416 > 0,1$);

- значения частного F-критерия для параметров уравнения, по сравнению с предыдущим случаем (рисунок 4.11а), уменьшились (рисунок 4.13а), а для частного коэффициента стоящего перед X^3 превышает критическое значение, значит введение дополнительного члена в уравнение (X^3) было не оправдано;

- коэффициенты корреляции и детерминации (рисунок 4.13в), по сравнению с предыдущей моделью (рисунок 4.11в), однако мера значимости линии регрессии увеличились уменьшилась;

- на диаграмме рассеивания заметных изменений не наблюдается (рисунок 4.13г), по сравнению с линейной моделью (рисунок 4.11г).

Все выше перечисленное говорит о том, что уравнение регрессии $Y = a_0 + a_1X + a_2X^2$ «лучше» описывает взаимосвязь X с Y чем уравнение $Y = a_0 + a_1X + a_2X^2 + a_3X^3$.

Можно сделать общее заключение, что взаимосвязь между X с Y описывается уравнением

$$Y = -0,34151 - 2,39701 * X - 2,40653 * X^2.$$

Корреляционно-регрессионный анализ в *LibreOffice*

Регрессионный анализ в *LibreOffice*

Рассмотрим на примере построение регрессионной модели в *LibreOffice* и интерпретацию результатов. Возьмем линейный тип регрессии.

Пример. На 6 предприятиях была проанализирована среднемесячная заработная плата и количество уволившихся сотрудников. Исходные данные на рисунке 4.14. Необходимо определить зависимость числа уволившихся сотрудников от средней зарплаты .

	A	B	C
1		К-во ув.	З/п
2		y	x
3	1	60	100
4	2	35	150
5	3	20	200
6	4	20	250
7	5	15	300
8	6	15	350

Рисунок 4.14 – Исходные данные

Модель линейной регрессии имеет следующий вид:

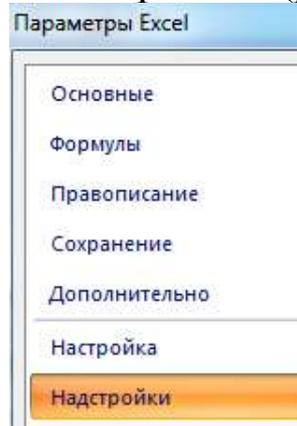
$$Y = a_0 + a_1x_1 + \dots + a_kx_k.$$

где a – коэффициенты регрессии, x – влияющие переменные, k – число факторов.

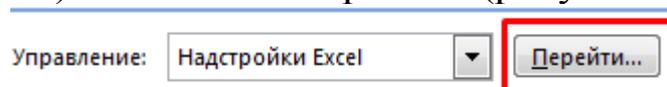
В нашем примере в качестве Y выступает показатель уволившихся работников. Влияющий фактор – заработная плата (x).

В *LibreOffice* существуют встроенные функции, с помощью которых можно рассчитать параметры модели линейной регрессии. Но быстрее это сделает надстройка «Пакет анализа».

Нажимаем кнопку «Офис» и переходим на вкладку «Параметры *LibreOffice*». «Надстройки» (рисунок 4.15).

Рисунок 4.15 – Вкладка «Параметры *LibreOffice*». «Надстройки»

Внизу, под выпадающим списком, в поле «Управление» будет надпись «Надстройки *LibreOffice*» (если ее нет, нажмите на флажок справа и выберите). И кнопка «Перейти» (рисунок 4.16).

Рисунок 4.16 – Надстройки *LibreOffice*

Открывается список доступных надстроек. Выбираем «Пакет анализа» (рисунок 4.17) и нажимаем ОК.

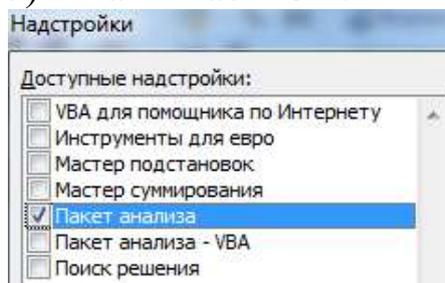


Рисунок 4.17 – Выбор «Пакет анализа»

После активации надстройка будет доступна на вкладке «Данные» (рисунок 4.18).

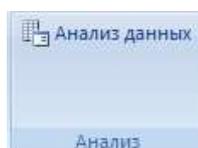


Рисунок 4.18 – Инструмент «Анализ данных»

Регрессионный анализ

Открываем меню инструмента «Анализ данных». Выбираем «Регрессия» (рисунок 4.19).

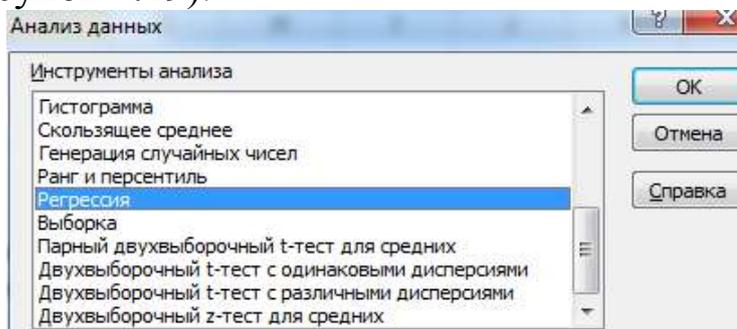


Рисунок 4.19 – Выбор инструмент анализа «Регрессия»

Откроется меню для выбора входных значений и параметров вывода (где отобразить результат). В полях для исходных данных указываем диапазон описываемого параметра (Y) и влияющего на него фактора (X). Остальное можно и не заполнять (рисунок 4.20).

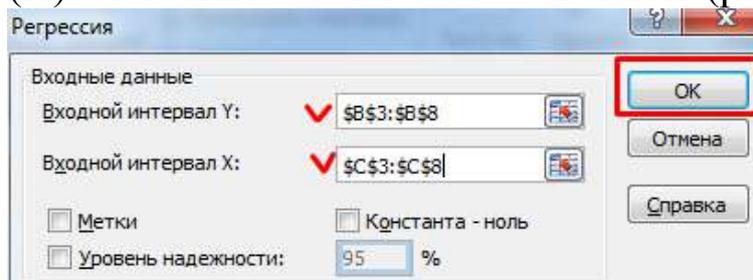


Рисунок 4.20 – Ввод исходных данных

После нажатия ОК, программа отобразит расчеты на новом листе (можно выбрать интервал для отображения на текущем листе или назначить вывод в новую книгу). Результаты расчётов показаны на рисунке 4.21.

	A	B	C	D	E	F	G	H	I
1	ВЫВОД ИТОГОВ								
3	<i>Регрессионная статистика</i>								
4	Множественный R	0,868736918							
5	R-квадрат	0,754703833							
6	Нормированный R-квадрат	0,693379791							
7	Стандартная ошибка	9,710083125							
8	Наблюдения	6							
10	<i>Дисперсионный анализ</i>								
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>ачимость F</i>			
12	Регрессия	1	1160,357143	1160,357143	12,30681818	0,024714			
13	Остаток	4	377,1428571	94,28571429					
14	Итого	5	1537,5						
16			<i>Стандарт</i>			<i>Нижние</i>	<i>Верхние</i>	<i>Нижние</i>	<i>Верхние</i>
		<i>Коэффициент</i>	<i>ная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>95%</i>	<i>95%</i>	<i>95,0%</i>	<i>95,0%</i>
17	Y-пересечение	64,14285714	11,17212274	5,741331225	0,004560379	33,12407	95,16164	33,12407	95,16164
18	Переменная X 1	-0,162857143	0,046423077	-3,508107493	0,024714164	-0,29175	-0,03397	-0,29175	-0,03397

Рисунок 4.21 – Результаты расчётов

В первую очередь обращаем внимание на R-квадрат и коэффициенты.

R-квадрат – коэффициент детерминации. В нашем примере – 0,755, или 75,5%. Это означает, что расчетные параметры модели на 75,5% объясняют зависимость между изучаемыми параметрами. Чем выше коэффициент детерминации, тем качественнее модель. Хорошо – выше 0,8. Плохо – меньше 0,5 (такой анализ вряд ли можно считать резонным). В нашем примере – «неплохо».

Коэффициент 64,1428 показывает, каким будет Y, если все переменные в рассматриваемой модели будут равны 0. То есть на значение анализируемого параметра влияют и другие факторы, не описанные в модели.

Коэффициент -0,16285 показывает весомость переменной X на Y. То есть среднемесячная заработная плата в пределах данной модели влияет на количество уволившихся с весом -0,16285 (это небольшая степень влияния). Знак «-» указывает на отрицательное влияние: чем больше зарплата, тем меньше уволившихся.

Корреляционный анализ в *LibreOffice*

Рассмотрим, как с помощью средств *LibreOffice* найти коэффициент корреляции.

Пример. Определить, есть ли взаимосвязь между временем работы токарного станка и стоимостью его обслуживания. Исходные данные для анализа представлены на рисунке 4.22.

	А	В
1	Время работы станка, тыс.ч.	Ст-сть обслужив., тыс.руб.
2	0,5	7,5
3	0,6	7,75
4	0,7	7,25
5	0,8	7,4
6	0,9	7,9
7	1	8
8	1,1	8,5
9	1,2	8,4
10	1,3	8,35
11	1,4	8,55
12	1,5	8,7
13	1,6	9,1
14	1,7	8,8

Рисунок 4.22 – Исходные данные

Ставим курсор в любую ячейку и нажимаем кнопку fx.

В категории «Статистические» выбираем функцию КОРРЕЛ.

Аргумент «Массив 1» - первый диапазон значений – время работы станка: A2:A14.

Аргумент «Массив 2» - второй диапазон значений – стоимость ремонта: B2:B14. Нажимаем ОК (рисунок 4.23).

fx =КОРРЕЛ(A2:A14;B2:B14)		
В	С	D
Ст-сть обслужив., тыс.руб.		
7,5	0,924771	

Рисунок 4.23 – определение коэффициента корреляции

Для корреляционного анализа нескольких параметров (более 2) удобнее применять «Анализ данных» (надстройка «Пакет анализа»). В списке нужно выбрать корреляцию и обозначить массив.

Полученные коэффициенты отобразятся в корреляционной матрице (рисунок 4.24).

	A	B	C
1		Столбец 1	Столбец 2
2	Столбец 1	1	
3	Столбец 2	0,92477117	1

Рисунок 4.24 – Корреляционная матрица

Построение диаграммы разброса и регрессионной модели в LibreOffice

Пример. Исходные данные для построения диаграммы разброса и регрессионной модели представлены на рисунке 4.25.

	A	B	C
1		Цена за ед., руб.	Объем продаж, тыс.руб.
2	n	X	Y
3	1	13	520
4	2	13	550
5	3	15	600
6	4	15	610
7	5	16	620
8	6	21	724
9	7	21	680
10	8	14	300
11	9	40	962
12	10	12	270

Рисунок 4.25 – Исходные данные для построения диаграммы разброса и регрессионной модели

Строим корреляционное поле: «Вставка» - «Диаграмма» - «Точечная диаграмма» (дает сравнивать пары). Диапазон значений – все числовые данные таблицы, представленной на рисунке 4.25. Полученная диаграмма разброса или поле корреляции представлена на рисунке 4.26.

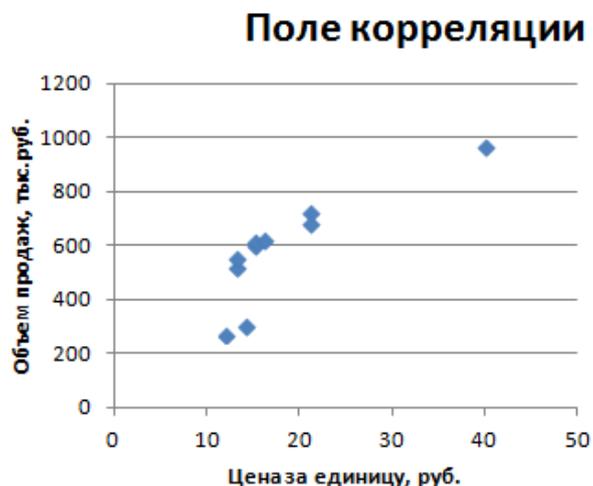


Рисунок 4.26 – Диаграмма разброса

Щелкаем левой кнопкой мыши по любой точке на диаграмме. Потом правой. В открывшемся меню выбираем «Добавить линию тренда» (рисунок 4.27).

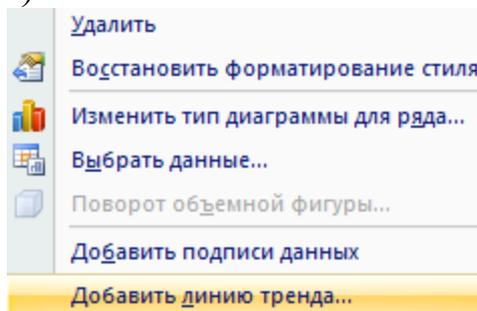


Рисунок 4.27 – Добавление линии тренда

Назначаем параметры для линии. Тип – «Линейная». Внизу – «Показать уравнение на диаграмме» (рисунок 4.28). Нажимаем «Заккрыть». Полученная регрессионная модель и данные регрессионного анализа представлена на рисунке 4.29.

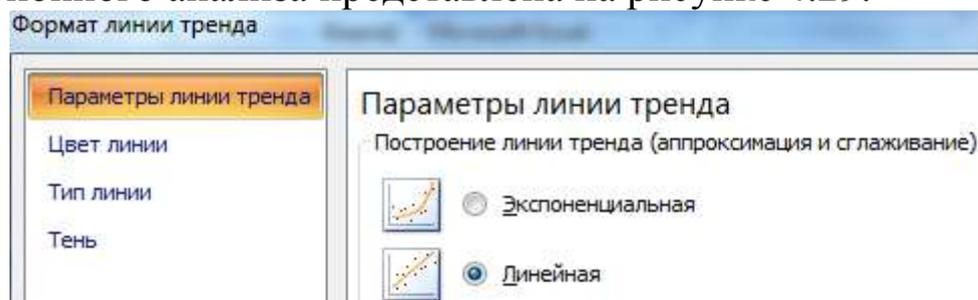


Рисунок 4.28 – Назначение параметров линии регрессии

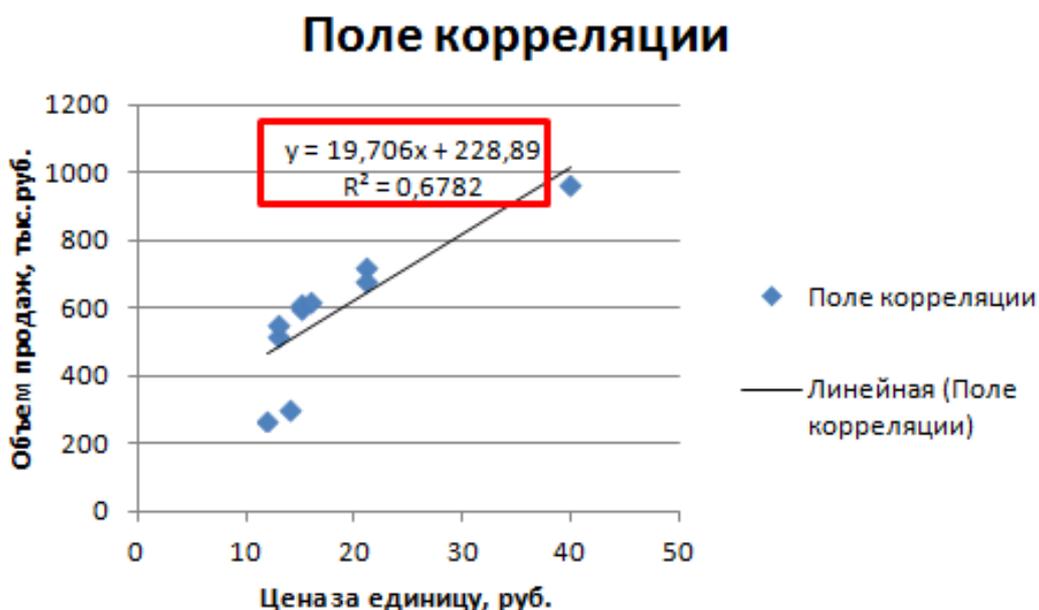


Рисунок 4.29 – Регрессионная модель и данные регрессионного анализа

Порядок выполнения работы

Получив у преподавателя исходные данные для выполнения лабораторной работы, студент изучает методику проведения корреляционного и регрессионного анализа. Выполняет корреляционный и регрессионный анализ на компьютере.

Содержание отчета

Отчет по лабораторной работе должен содержать следующие пункты:

- название лабораторной работы;
- цель работы;
- краткие теоретические сведения;
- краткое описание хода выполнения работы;
- индивидуальное задание для выполнения лабораторной работы;
- результаты выполнения работы: таблица с исходными данными, диаграмма разброса (рассеяния), таблица корреляционной матрицы, таблицы параметров уравнения регрессии.
- выводы об исследуемой зависимости рассмотренных параметров процесса и проверке значимости корреляции и регрессии.

Контрольные вопросы

1. Что такое «диаграмма разброса»?
2. Назначение и применение диаграммы разброса.
3. Каковы правила построения диаграммы разброса?
4. Что называется «корреляционной зависимостью»?
5. Виды корреляционной зависимости и их связь с коэффициентом корреляции.
6. Что характеризует коэффициент корреляции?
7. Что такое «ковариация»?
8. Что называется «регрессионным анализом»?
9. Что рассчитывается при построении регрессионного уравнения?
10. Какова интерпретация коэффициентов уравнения регрессии?
11. Что такое линейная регрессия?

12. Что такое многофакторная регрессия?
13. Что такое нелинейная регрессия?
14. Что рассчитывается при построении регрессионного уравнения?
15. Что такое критерий значимости регрессии?
16. Что такое коэффициент детерминации?
17. Какова интерпретация коэффициентов уравнения множественной регрессии?
18. Как осуществляется выбор «наилучшего» уравнения регрессии?

Лабораторная работа №5

Применение контрольных карт Шухарта для анализа стабильности и управления технологическим процессом

Цель работы: получить сведения о назначении, типах и методах построения контрольных карт по количественному признаку и получить практические навыки обработки статистических данных и построения контрольных карт по количественному признаку в программе *STATISTICA* или *LibreOffice* для анализа стабильности и управления технологическим процессом.

Задание для самостоятельного выполнения

Построить контрольную карту по количественному признаку для анализа точности, стабильности и воспроизводимости технологического процесса по данным контроля двух размеров полученных в процессе изготовления выпускаемого изделия.

Краткие теоретические сведения

Понятие стабильности и управляемости технологического процесса

Статистическое управление процессами состоит в выявлении неслучайных нарушений процесса; при этом управляющее воздействие (регулирование) применяется тогда, когда выпускаемая продукция или услуга еще удовлетворяют заданным требованиям, но некоторые статистические показатели дают основание предполагать наличие неслучайной причины, которая приведет к нарушению процесса.

Изменение характеристик качества изделий в процессе изготовления обусловлено причинами двух типов. Одна группа причин связана с особенностями данного процесса – износом инструмента, ослаблением крепления, изменением температуры охлаждающей жидкости и т.д. Это неслучайные причины вариаций, которые могут быть устранены при настройке процесса. Другая группа причин - неустраняемые, случайные причины изменчивости (колебания температуры окружающей среды, вариации характеристик материала и т.п.).

Технологический процесс желательно проводить так, чтобы изменчивость характеристик качества была обусловлена только

случайными причинами. Неслучайные причины изменчивости процесса могут быть выявлены с помощью статистических методов. управление технологическим процессом состоит в выявлении и устранении этих причин. Изменчивость за счет случайных причин может быть снижена только путем усовершенствования самого процесса.

Вмешательство в технологический процесс для настройки требуется тогда, когда выпускаемая продукция еще удовлетворяет техническим требованиям, но статистические показатели процесса свидетельствуют о наличии неслучайных воздействий.

Процесс считается статистически управляемым, если обеспечена его стабильность, т.е. повторяемость контролируемых параметров. Статистическая управляемость еще не означает, что процесс удовлетворяет требованиям потребителя: стабильный процесс может не обеспечивать выполнение требуемых технических условий. В связи с этим, кроме оценки стабильности процесса, проверяются и его возможности: как соотносятся характеристики процесса с техническими условиями.

Типы контрольных карт по количественному признаку

Инструментом оперативного статистического управления процессом является контрольная карта, предложенная У. Шухартом: сигнал о разладке процесса подается при выходе контролируемого показателя за некоторую пороговую границу. Контрольная карта Шухарта – графическое изображение мониторинга процесса.

Карты по количественному признаку применяются тогда, когда контролируемый показатель можно измерить. Это может быть, например, диаметр вала, шероховатость поверхности, крутящий момент, прочность изделия, температура нагрева, электрическое сопротивление, вес изделия, влажность, содержание некоторого вещества в растворе и т.д.

Обычно контролируется изменение как среднего значения показателя качества, характеризующего уровень настройки процесса, так и технологического рассеивания: строят двойные карты Шухарта. При этом уровень настройки процесса может оцениваться по средним значениям или медианам, а рассеивание –

по стандартным отклонениям или размахам. В соответствии с этим чаще всего используются двойные карты следующих типов:

- карты средних значений и размахов ($\bar{X} - R$ - карты);
- карты средних значений и стандартных отклонений ($\bar{X} - s$ - карты);
- карты медиан и размахов ($Me - R$ - карты).

Процесс считается стабильным, или статистически управляемым, если об этом свидетельствуют обе карты – и для среднего уровня, и для рассеивания.

Построению контрольной карты должно предшествовать предварительное исследование технологического процесса.

В ходе исследования:

- устанавливается связь между процессом и качеством продукции для выработки рекомендаций по наладке процесса;
- если анализ показывает, что процесс находится в неуправляемом состоянии, и возможности процесса не соответствуют установленным требованиям, то необходимо определить причины сдвигов уровня процесса и способы настройки.

Кроме этого необходимо:

- убедиться, что выборочные средние арифметические \bar{X} распределены по нормальному закону;
- убедиться, что индивидуальные измерения, используемые для вычисления \bar{X} , проводятся измерительным инструментом, деление шкалы которого не превышает $\sigma/2$.

Для предварительного исследования технологического процесса на контроль отбирают, как минимум, $N = 100$ единиц продукции.

При этом правильнее отбирать единицы продукции не подряд, а, например, каждую пятую, десятую и т.п., чтобы охватить, возможно, больший интервал времени работы оборудования.

У отобранных единиц продукции контролируют заданный параметр X . Объем выборки n устанавливают не менее трёх и не более десяти единиц продукции.

Чтение контрольных карт

Зоны А, В, С

Обычно для задания критериев поиска серий область

контрольной карты над центральной линией и под ней делится на три «зоны» (рисунок 5.1).

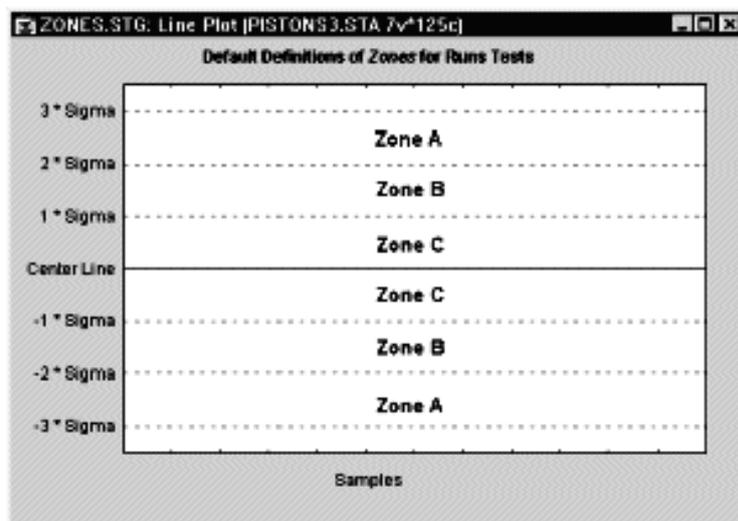


Рисунок 5.1 – Зоны А, В, С

По умолчанию, зона А определяется как область, расположенная на расстоянии от 2 до 3 сигма по обе стороны от центральной линии. Зона В определяется как область, отстоящая от центральной линии на расстояние от 1 до 2 сигма, а зона С – как область, расположенная между центральной линией по обе ее стороны и ограниченная прямой, проведенной на расстоянии одной сигма от центральной линии.

9 точек в зоне С или за ее пределами (с одной стороны от центральной линии)

Если этот критерий выполняется (т.е. если на контрольной карте обнаружено такое расположение точек), то делается вывод о возможном изменении среднего значения процесса в целом. Заметим, что здесь делается предположение о симметричности распределения исследуемых характеристик качества вокруг среднего значения процесса на графике. Но это условие не выполняется, например, для R-карт, S-карт и большинства карт по альтернативному признаку. Тем не менее, данный критерий полезен для того, чтобы указать занимающемуся контролем качества инженеру на присутствие потенциальных трендов процесса. Например, здесь стоит обратить внимание на последовательные выборочные значения с изменчивостью ниже среднего, так как с их помощью можно догадаться, каким образом снизить вариацию процесса.

6 точек монотонного роста или снижения, расположенные подряд

Выполнение этого критерия сигнализирует о сдвиге среднего значения процесса. Часто такой сдвиг обусловлен изнашиванием инструмента, ухудшением технического обслуживания оборудования, повышением квалификации рабочего и т.п. (Nelson, 1985).

14 точек подряд в "шахматном" порядке (через одну над и под центральной линией)

Если этот критерий выполняется, то это указывает на действие двух систематически изменяющихся причин, которое приводит к получению различных результатов. Например, в данном случае может иметь место использование двух альтернативных поставщиков продукции или отслеживание двух различных альтернативных воздействий.

2 из 3-х расположенных подряд точек попадают в зону А или выходят за ее пределы

Этот критерий служит "ранним предупреждением" о начинающейся разладке процесса. Заметим, что для данного критерия вероятность получения ошибочного решения (критерий выполняется, однако процесс находится в нормальном режиме) в случае \bar{X} -карт составляет приблизительно 2 %.

4 из 5-ти расположенных подряд точек попадают в зону В или за ее пределы

Как и предыдущий, этот критерий может рассматриваться в качестве индикатора - "раннего предупреждения" о возможной разладке процесса. Процент принятия ошибочного решения о наличии разладки процесса для этого критерия также находится на уровне около 2%.

15 точек подряд попадают в зону С (по обе стороны от центральной линии)

Выполнение этого критерия указывает на более низкую изменчивость по сравнению с ожидаемой (на основании выбранных контрольных пределов).

8 точек подряд попадают в зоны В, А или выходят за контрольные пределы, по обе стороны от центральной линии (без попадания в зону С)

Выполнение этого критерия служит свидетельством того, что различные выборки подвержены влиянию различных факторов, в результате чего выборочные средние значения оказываются распределенными по бимодальному закону. Такая ситуация может сложиться, например, когда отмечаемые на \bar{X} -карте выборки изделий были произведены двумя различными станками, один из которых производит изделия со значением контролируемой характеристики выше среднего, а другой - ниже.

Анализ качества технологического процесса

Уровень качества процесса связан с анализом его реальных и возможных характеристик.

Предварительно следует сказать о взаимосвязи требований технических условий (ТУ) к некоторому процессу или изделию с величинами граничных допусков на контрольных картах. Может возникнуть естественное желание использовать в качестве граничных допусков на контрольных картах значения, взятые из ТУ. Такой подход будет ошибочным по двум причинам. Во-первых, допуски на контрольных картах должны базироваться на изменчивости реального производственного процесса, а требования ТУ формируются разработчиками обычно до начала производства и совсем из других соображений. Отсюда следует, что между требованиями ТУ и граничными зонами на контрольной карте нет прямой связи. Вторая причина связана с тем, что, например, на часто применяемые контрольные карты средних значений наносят осредненные значения контролируемого параметра, а требования ТУ относятся к каждому отдельному образцу. Естественно при этом попытка сопоставления осредненных и индивидуальных характеристик будет ошибочной.

Показатель качества изделия должен находиться в пределах поля допуска. Для определения того, способен ли технологический процесс выпускать изделия, имеющие показатели качества в пределах допуска, используются индексы воспроизводимости. Использование этих коэффициентов считается более удобным и наглядным, нежели использование таких характеристик, как среднее значение и среднее стандартное отклонение.

Индекс воспроизводимости – безразмерная величина, показывающая связь между характеристиками технологического

процесса и допуском. Пусть U и L – соответственно верхняя и нижняя границы поля допуска, а σ – среднеквадратическое отклонение показателя качества в технологическом процессе. Тогда, если показатель имеет нормальное распределение и его среднее значение находится в середине поля допуска, индекс воспроизводимости C_p определяется по формуле:

$$C_p = \frac{U - L}{6\sigma} \quad (5.1)$$

где при контроле технологического процесса с помощью контрольных карт Шухарта в качестве несмещенной оценки стандартного отклонения применяется величина

$$\sigma = \bar{s}/c \quad (5.2)$$

если расчет ведется через стандартные отклонения, или

$$\sigma = \bar{R}/d \quad (5.3)$$

если используются размахи.

Величины коэффициентов c и d определяются из таблицы 5.1 в зависимости от объема выборки.

Таблица 5.1

Значения коэффициентов c и d

n	2	3	4	5	6	7	8	9	10
c	0,798	0,889	0,921	0,940	0,952	0,959	0,965	0,969	0,973
d	1,128	1,1693	2,059	2,326	2,534	2,704	2,847	2,970	3,078

Воспроизводимость технологического процесса оценивают, исходя из следующих критериев:

1. В случае, когда $C_p \geq 1,67$, ширина интервала между контрольными нормативами не менее чем в 10 раз превышает стандартное отклонение s ; разброс параметров изделия невелик, появление брака не угрожает. При $C_p = 1,67$ уровень несоответствий составляет 6 изделий на 1 млн. выпущенной продукции, а при $C_p = 2$, когда поле допуска вдвое шире диапазона рассеивания технологического процесса, - 2 изделия на миллион.

2. В случае, когда $1,67 > C_p \geq 1,33$, ширина интервала между контрольными нормативами в 8-10 раз превышает стандартное отклонение s . Стандарт рекомендует в качестве минимального приемлемого значения $C_p = 1,33$, при этом процент несоответствующих единиц продукции составляет 0,0066%.

3. В случае, когда $1,33 > C_p \geq 1,00$, ширина интервала между контрольными нормативами в 6-8 раз превышает стандартное отклонение s . Когда показатель $C_p = 1$, вероятность появления брака составляет 0,27%, поэтому необходимо усилить контроль процесса, провести анализ факторов, влияющих на разброс, и провести мероприятия по улучшению состояния процесса.

4. В случае, когда $1,00 > C_p \geq 0,67$, ширина интервала между нижней и верхней границами нормы всего лишь в 4-6 раз превышает стандартное отклонение s . Когда показатель C_p приближается к 0,67, вероятность появления брака составляет 4,56%. Это означает, что контроль процесса не удовлетворителен. Необходимо наладить строгий контроль процесса и провести сплошной контроль выпускаемых изделий с целью недопущения брака. Вместе с тем нужно провести немедленно исследование факторов, влияющих на разброс, и принять меры к улучшению состояния процесса.

5. В случае, когда $0,67 > C_p$, ширина интервала между нижней и верхней границами нормы не превышает $4s$. Процент брака превышает 4,56%. О таком процессе можно сказать, что он неконтролируем. Необходимо провести сплошной контроль продукции, чтобы предотвратить выпуск бракованных изделий, и одновременно принять меры к повышению качества, выявить причины появления брака. Иногда приходится заново проводить изучение потребностей потребителей, а также пересматривать нормы.

Часто используется обратная величина C_p , называемая коэффициентом точности:

$$CR = 1/C_p \quad (5.4)$$

Он вычисляется в процентах и показывает часть поля допуска, покрываемую рассеиванием процесса. Например, $C_p = 1,33$ соответствует 75% поля допуска.

При одностороннем допуске вместо формулы (5.1) используют соответственно верхний индекс воспроизводимости

$$C_{pkU} = \frac{U - \bar{x}}{3\sigma} \quad (5.5)$$

где \bar{x} – средний уровень настройки процесса,

или нижний индекс воспроизводимости

$$C_{PKL} = \frac{\bar{x} - L}{3\sigma} \quad (5.6)$$

Индекс воспроизводимости C_p предполагает точное центрирование процесса – совпадение среднего значения \bar{x} с целевым уровнем μ . Для учета расхождения между этими характеристиками вводится индекс центрированности k :

$$k = 2|\mu - \bar{x}| / (U - L) \quad (5.7)$$

при точном центрировании $k = 0$, при совпадении среднего уровня с одной из границ поля допуска $k = 1$.

Реальная воспроизводимость процесса с учетом центрирования определяется индексом настройки или налаженности

$$C_{pk} = C_p(1 - k) \quad (5.8)$$

И не превышает индекса воспроизводимости. Если $C_p = C_{pk}$, то это говорит о том, что процесс отцентрован.

Индекс настройки может, определен и по формуле:

$$C_{pk} = \min(C_{PKU}, C_{PKL}) \quad (5.9)$$

Данный индекс является количественной характеристикой степени соответствия технологического процесса требованиям заказчика.

Параметр C_{pk} называют также коэффициентом стабильности и воспроизводимости технологического процесса. Этот индекс можно интерпретировать как запас точности технологического процесса по отношению к требованиям заказчика.

В приведенной формуле предполагается, что среднее значение лежит внутри поля допуска; в противном случае полагают, что $C_{pk} = 0$.

Если $C_{pk} < 1$, то технологический процесс не может считаться удовлетворительным для нужд серийного производства.

Наиболее типичное требование состоит в том, чтобы технологический процесс удовлетворял условию $C_{pk} > 1,33$. Ситуация, когда $C_{pk} \geq 2$, обычно считается близкой к идеалу.

Коэффициент C_{pk} характеризует возможности технологического процесса. Он позволяет также отследить динамику (тренд) процесса во времени.

Зная значения индекса воспроизводимости и индекса настройки процесса можно определить уровень несоответствий продукции по таблице, представленной в Приложении Е.

Рассмотренные показатели применимы для оценки возможностей процесса в том случае, если с использованием контрольных карт подтверждена статистическая управляемость процесса.

Коэффициент годности является частным случаем коэффициента точности, который применяется при анализе воспроизводимости процесса по критериям точности и стабильности и определяется как

$$C_m = T/(k_p s), \quad (5.10)$$

где k_p – коэффициент, зависящий от типа закона распределения исследуемых данных (для нормального закона распределения $k_p = 6$, для закона равной вероятности $k_p = 3,464$).

Еще одним основным параметром, характеризующим качество процесса, является степень засоренности данных g , равная количеству резко выделяющихся наблюдений в процентах:

$$g = \frac{k_v}{n} \cdot 100\% \quad (5.11)$$

где k_v – количество выделяющихся наблюдений (выбросов); n – полное число наблюдений (объем выборки).

Чаще всего появление резко выделяющихся наблюдений связано со сбоем измерительного оборудования, а также с несовершенством технологических образцов и методик контроля. Устранение причин появления резко выделяющихся наблюдений требует согласованных действий различных служб: технологических, метрологических, обслуживающих оборудование, поставляющих материалы, обеспечивающих состояние производственных сред и других.

Коэффициенты C_{pk} и g характеризуют качество процесса с различных сторон, дополняя друг друга. В то время, как первый из них характеризует возможности процесса, не учитывая различного рода обычно редких «досадных» промахов и ошибок, второй, напротив, как раз и характеризует частоту появления таких сбоев в технологии и измерениях.

С методической и технической точек зрения недопустимо

игнорировать коэффициент засоренности и производить расчет коэффициента C_{pk} по засоренным данным без их фильтрации от резко выделяющихся наблюдений.

Порядок построения контрольных карт по количественному признаку в программе STATISTICA

Рассмотрим пример построения $\bar{X} - R$ контрольной карты по данным результата контроля 100 диаметров колец подшипников после процесса шлифовки. Построим контрольные карты по количественному признаку в программе STATISTICA.

Вводим исходные данные

Запускаем программу STATISTICA, создаём новую таблицу, состоящую из одного столбца и ста строк. Вводим имеющиеся данные в ячейки таблицы.

Строим контрольную карту

Для построения контрольных карт открываем следующие вкладки: **Индустриальная статистика и Сигма шесть** → **Качество диаграммы управления** (рисунок 5.2).

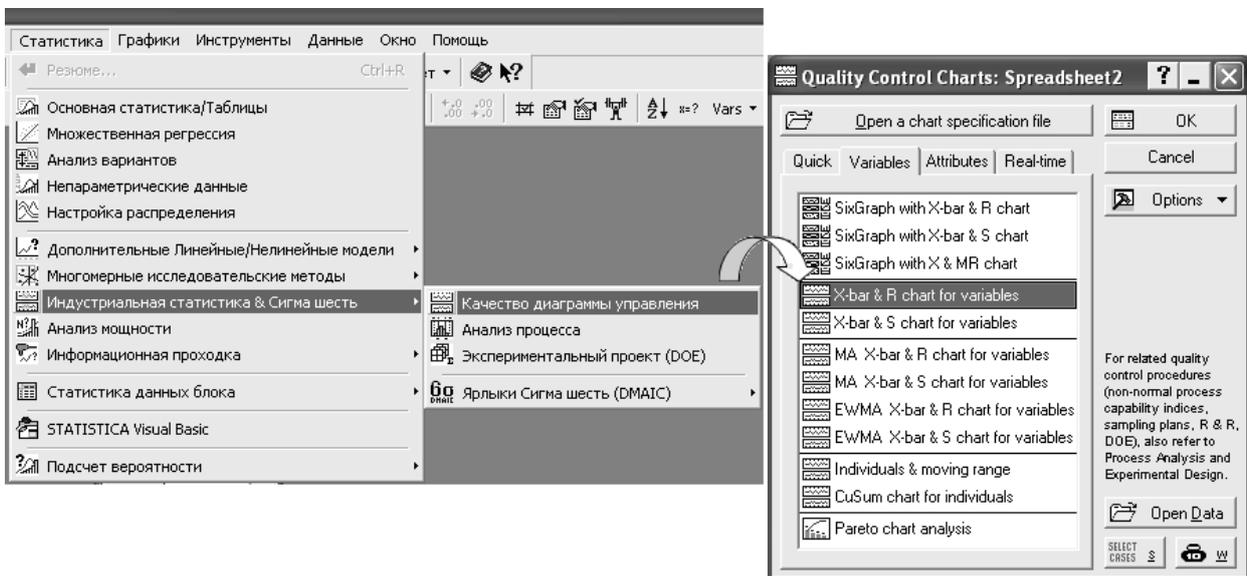


Рисунок 5.2 – Вкладка контрольных карт

В следующем окне выбираем вкладку «Variables» и далее строку «X-bar & R chart for variables», и нажимаем кнопку «ОК». В окне «Defining Variables for X-bar and R Chart» во вкладке «Quick» в строке «Constant sample size» ставим галочку и указываем объем выборки, равный 5, кнопкой «Variables» указываем номер столбца с данными. В поле «Measurements (observations):» должно

появиться имя выбранной переменной. После нажатия кнопки «ОК» программа строит контрольные X/R – карты.

Настройка контрольных карт

Полученная карта не удовлетворяет нашим требованиям, т.к. на ней не показаны зоны А, В и С. Чтобы их обозначить на X-карте открываем меню в нижней левой части окна программы «X-bar/R:…» (необходимо нажать на соответствующую надпись). В диалоговом окне «X-Bar/R» (рисунок 5.3) выбираем вкладку «X (MA...) specs» и нажимаем кнопку «Warning lines», в появившемся окошке «Warning lines» ставим маркер на строке «Multiples of sigma» (Кратно сигма), в строке «Lower» (нижний) указываем «-2», в строке «Upper» (верхний) указываем «2», это значит, что программа выделит линиями двухсигмовую зону ниже центральной линии и двухсигмовую зону выше центральной линии на X-карте.

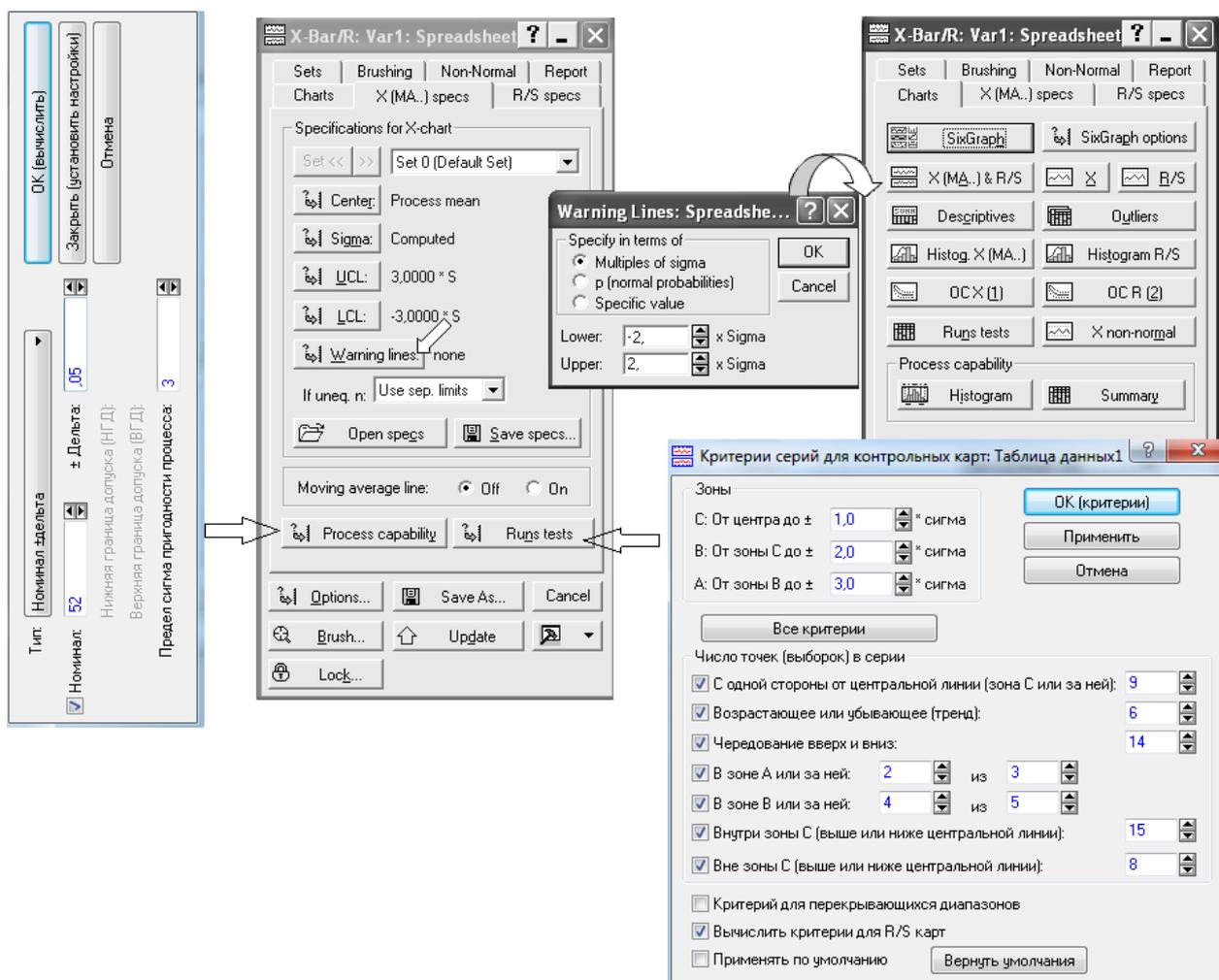


Рисунок 5.3 – Меню построения контрольных карт

Нажимаем кнопку «ОК» и автоматически возвращаемся к предыдущему окну. Аналогично настраиваем контрольные зоны на R-карте (вкладка «R/S specs»). После проведенных настроек выбираем вкладку «Charts» и нажимаем кнопку «X (MA...) & R/S» (см. рисунок 5.3) и программа строит X/R контрольную карту (рисунок 5.4).

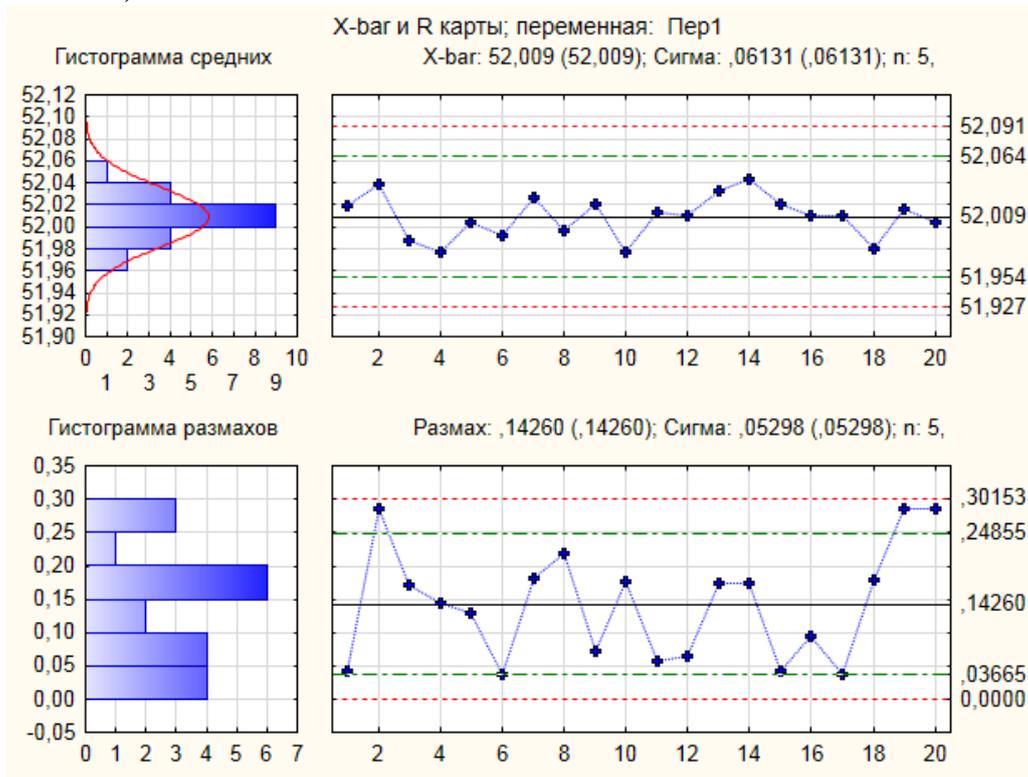


Рисунок 5.4 – X/R контрольная карта

Чтение контрольных карт и оценка стабильности процесса

Чтобы провести тесты (чтение) контрольных карт по приведенным выше признакам, в диалоговом окне «X-Bar/R» (рисунок 5.3) выбираем вкладку «X (MA...) specs» и нажимаем кнопку «Run tests» и проводим настройку критериев серий для контрольной карты, и нажимаем кнопку «Применить».

Открываем окно «X-bar/R:...» и во вкладке «Charts» нажать кнопку «Run tests» (рисунок 5.3). Программа строит таблицу с результатами теста для X – карты (рисунок 5.5) и R – карты (рисунок 5.6).

На рисунках 5.5 и 5.6 обозначено:

9 samples on same side of center – 9 точек в зоне С или за её пределами (с одной стороны от центральной линии).

Зоны A/B/C: 3,000/2,000/1,000 * Сигма Критерии против альтернатив спец. вида	Пер1 ; Критерии серий (Таблица данных1) X-bar карта Центр: 52,009010 Сигма: 0,027418	
	от выборки	до выборки
9 точек по одну сторону от центра	ОК	ОК
6 точек в возр./уменьш. ряду	ОК	ОК
14 точек в "шахматном порядке"	ОК	ОК
2 из 3 точек в зоне А или вне ее	ОК	ОК
4 из 5 точек в зоне В или вне ее	ОК	ОК
15 точек в зоне С	ОК	ОК
8 точек вне зоны С	ОК	ОК

Рисунок 5.5 – Таблица результатов теста X – карты

Зоны A/B/C: 3,000/2,000/1,000 * Сигма Критерии против альтернатив спец. вида	Пер1 ; Критерии серий (Таблица данных1) R карта Центр: 0,142600 Сигма: 0,052976	
	от выборки	до выборки
9 точек по одну сторону от центра	ОК	ОК
6 точек в возр./уменьш. ряду	ОК	ОК
14 точек в "шахматном порядке"	ОК	ОК
2 из 3 точек в зоне А или вне ее	18	20
4 из 5 точек в зоне В или вне ее	ОК	ОК
15 точек в зоне С	ОК	ОК
8 точек вне зоны С	ОК	ОК

Рисунок 5.6 – Таблица результатов теста R – карты

6 samples in row in/decreasing – 6 точек монотонного роста или снижения, расположенные подряд.

14 samples alternating up & down – 14 точек подряд в «шахматном» порядке.

2 of 3 samples in Zone A or beyond - 2 из 3-х расположенных подряд точек попадают в зону А или выходят за ее пределы.

4 of 5 samples in Zone B or beyond - 4 из 5-ти расположенных подряд точек попадают в зону В или за ее пределы.

15 samples in Zone C - 15 точек подряд попадают в зону С (по обе стороны от центральной линии).

8 samples beyond Zone C - 8 точек подряд попадают в зоны В, А или выходят за контрольные пределы, по обе стороны от центральной линии (без попадания в зону С).

From sample to sample – от точки до точки.

Из рисунка 5.5 видно, что на X – карте отсутствуют сигнальные признаки (критерии особых причин), что говорит о стабильности технологического процесса по среднему уровню.

На рисунке 5.6 видно появление сигнального признака «2 of 3

samples in Zone A or beyond» на R-карте с 18 по 20 точку 2 из 3-х расположенных подряд точек попадают в зону А, что говорит о нестабильности размахов.

Таким образом, стабильность технологический процесс не подтверждена, и процесс признается статистически неуправляемым.

Оценка точности и воспроизводимости процесса

Для настройки анализа пригодности процесса в диалоговом окне «X-Bar/R» (рисунок 5.3) выбираем вкладку «X (MA...) specs» и нажимаем кнопку «Process capability» (Пригодность процесса) и указываем границы поля допуска, предварительно выбрав тип: номинал \pm дельта; нижний, номинал, верхний; нижний, номинал; верхний номинал. В рассматриваемом примере выбран тип номинал \pm дельта и указан допуск $52 \pm 0,05$ (см. рисунок 5.3). Нажимаем кнопку «ОК» и программа строит таблицы индексов производительности (рисунок 5.7) и пригодности процесса (рисунок 5.8).

		Пер1; Set 0 (Default Set)	
		-3,000 *Сигма= 51,82662	
		3,000 *Сигма= 52,19140	
Индексы производительности		Значен.	
Нижняя граница допуска		51,95000	
Номинал (плановые спецификации)		52,00000	
Верхняя граница допуска		52,05000	
PP (индекс производительности)		0,27414	
PR (отношение производительности)		3,64772	
PPK (подтвержд. производительность)		0,22474	
PPL (нижний индекс производительности)		0,32354	
PPU (верхний индекс производительности)		0,22474	

Рисунок 7 – Индексы производительности

		Пер1; Set 0 (Default Set)	
		-3,000 *Сигма= 51,82508	
		3,000 *Сигма= 52,19294	
Индекс пригодности процесса		Значен.	
Сигма внутри выборки=Rкарт./d2			
Нижняя граница допуска		51,95000	
Номинал (плановые спецификации)		52,00000	
Верхняя граница допуска		52,05000	
CP (потенциальная пригодность)		0,27185	
CR (отношение пригодности)		3,67853	
CPK (подтвержденное качество)		0,22286	
CPL (нижний индекс пригодности)		0,32083	
CPU (верхний индекс пригодности)		0,22286	
K (поправка на нецентрированность)		0,18020	

Рисунок 8 – Индексы пригодности процесса

Открываем окно «X-bar/R:…» и во вкладке «Charts» нажать кнопку «Histogram» (гистограмма) (рисунок 5.3). Программа строит гистограмму в пределах поля допуска и границ регулирования процесса (рисунок 5.9).

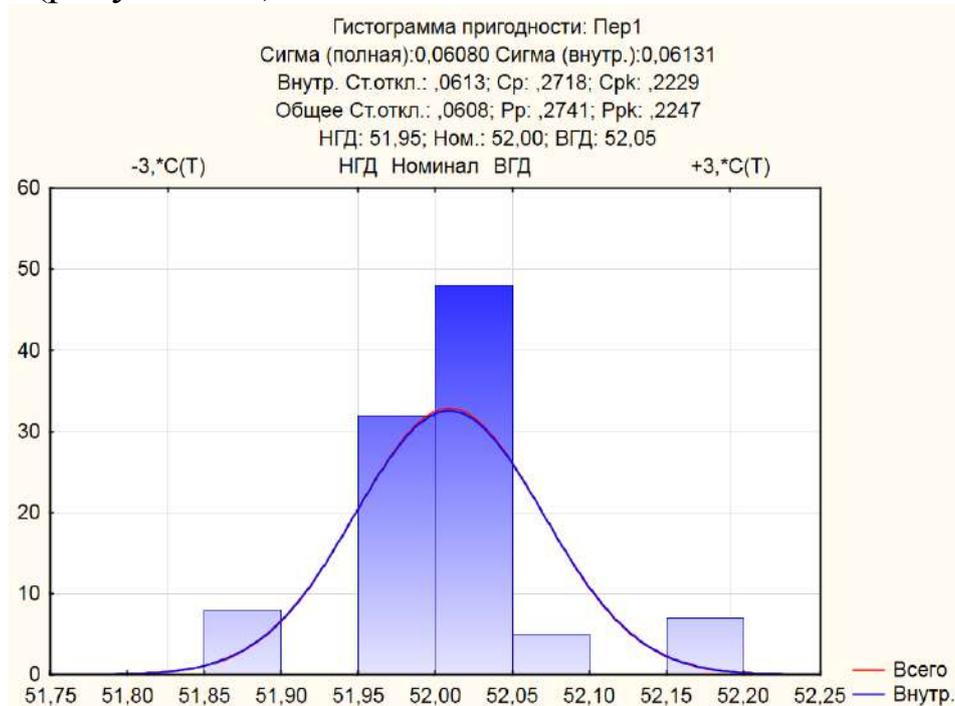


Рисунок 5.9 – Гистограмма

Построение контрольных карт в программе *LibreOffice*

В цехе принято решение перевести на статистическое регулирование технологический процесс изготовления болта на автоматах. За показатель качества выбран диаметр болта, равный 26 мм, и его допускаемые отклонения: $es = -0,005$ мм; $ei = -0,019$ мм. Построить контрольную $\bar{x} - s$ – карту и провести по ней статистический анализ процесса. Для упрощения измерений и вычислений измерительный прибор (рычажная скоба) был настроен на размер 25,980 мм. Результаты измерений (отклонения от размера 25,980 мм в микрометрах) приведены в таблице 5.2.

В ячейку A1 новой книги *LibreOffice* вводим заголовок работы. В диапазон A4:F24 вводим исходные данные (номера выборок и результаты контроля).

Вначале рассчитываем данные для построения контрольной карты средних значений. В ячейке G5 рассчитываем среднее значение первой выборки при помощи статистической функции СРЗНАЧ. Полученную формулу копируем в диапазон G6:G24.

Таблица 5.2

Результаты измерений отклонения от размера 25,980 мм в мкм

Цех автоматный		Оборудование – токарный автомат 5803		Контролируемая операция – нарезание резьбы		Контролируемый параметр – -0,005 Ø 26 -0,019	
Объём контроля N=100		Объём выборки n=5		Средство контроля – рычажная скоба			
Время	№ выборки	Результаты контроля					
7.00	1	10	3	5	14	10	
8.00	2	2	14	8	13	11	
9.00	3	12	12	3	8	10	
10.00	4	12	14	7	11	9	
11.00	5	10	11	9	15	7	
12.00	6	11	12	11	14	12	
13.00	7	15	11	14	8	3	
14.00	8	12	14	12	11	11	
15.00	9	11	7	11	13	9	
16.00	10	14	10	9	12	8	
7.00	11	9	11	14	10	13	
8.00	12	13	13	6	4	13	
9.00	13	5	8	3	3	4	
10.00	14	8	5	6	9	13	
11.00	15	8	4	9	5	8	
12.00	16	4	12	10	6	10	
13.00	17	10	6	13	10	5	
14.00	18	7	9	12	1	7	
15.00	19	4	7	6	7	12	
16.00	20	10	10	6	9	3	

В ячейке H5 рассчитываем значение \bar{x} (среднюю линию) как среднее из средних значений выборок при помощи статистической функции СРЗНАЧ. В полученной формуле для диапазона ячеек вводим абсолютную адресацию и копируем формулу в диапазон H6:H24. Это необходимо для того, чтобы в дальнейшем можно было провести среднюю линию на контрольной карте.

В ячейке В26 рассчитываем среднее квадратичное отклонение всей совокупности результатов измерений σ при помощи статистической функции СТАНДОТКЛОН для диапазона В5:F24.

В ячейке I5 рассчитываем нижнюю контрольную границу K_n . Формула в ячейке будет выглядеть так: =Н5-3*В26/КОРЕНЬ(5). Указав абсолютную адресацию для имён ячеек, копируем формулу из ячейки I5 в диапазон I6:I24. Это необходимо, чтобы в дальнейшем провести границу на карте.

В ячейке J5 рассчитываем верхнюю контрольную границу, и после указания абсолютной адресации для имён ячеек копируем формулу из ячейки J5 в диапазон J6:J24.

В ячейках K5 и L5 рассчитываем значения нижнего и верхнего технических допусков, вводя в них формулы =26000-19-25980 и =26000+5-25980 соответственно. Эти формулы копируем также в диапазон K6:L24.

Далее рассчитываем данные для построения контрольной карты средних квадратичных отклонений. В ячейке M5 рассчитываем среднее квадратичное отклонение первой выборки и копируем полученную формулу в диапазон M6:M24. В ячейке N5 рассчитываем среднее из СКО выборок, и после указания абсолютной адресации копируем формулу в диапазон N6:N24. В ячейке O5 рассчитываем нижнюю контрольную границу по формуле =N5*КОРЕНЬ(ХИ2ОБР(1-0,0027/2;4)/5) и копируем формулу в диапазон O6:O24. В ячейке P5 рассчитываем верхнюю контрольную границу и копируем содержимое ячейки в диапазон P6:P24.

Полученная электронная таблица показана на рисунке 5.10. По расчётным значениям строим $\bar{x} - s$ – карту.

	А	В	С	Д	Е	Ф	Г	Н	І	Ј	К	Л	М	Н	О	Р	
1	Лаб. работа 8. Контрольные карты по количественному признаку																
2																	
3							Карта средних значений						Карта СКО				
4	№ выборки	Результаты контроля				Хср	Хср ср	Кн	Кв	Тн	Тв	S	Scp	Кн	Кв		
5	1	10	3	5	14	10	8,4	9,15	4,5	13,8	1	15	4,39	3,05	0,44	5,76	
6	2	2	14	8	13	11	9,6	9,15	4,5	13,8	1	15	4,83	3,05	0,44	5,76	
7	3	12	12	3	8	10	9	9,15	4,5	13,8	1	15	3,74	3,05	0,44	5,76	

Рисунок 5.10 – Расчёт контрольных границ

Сначала строим $\bar{x} - s$ – карту. В мастере диаграмм выбираем вид

диаграммы **Точечная диаграмма**, на которой значения соединены отрезками. В качестве исходных данных выделяем диапазон A5:A24, G5:L24. Полученную диаграмму редактируем при помощи контекстного меню, а также наносим обозначения контрольных границ при помощи инструмента **Надпись** панели инструментов **Рисование**.

Аналогичным образом строим *s*-карту.

Чтобы получить из двух построенных карт единый объект, совмещаем их по длине (например, прижав к левому краю электронной таблицы), одновременно выделяем щелчками левой кнопкой мыши на каждой диаграмме при нажатой клавише Shift и группируем командой **Группировать**, вызываемой из инструмента **Действия** панели инструментов **Рисование**. Полученная контрольная $\bar{x} - s$ – карта показана на рисунке 5.11.

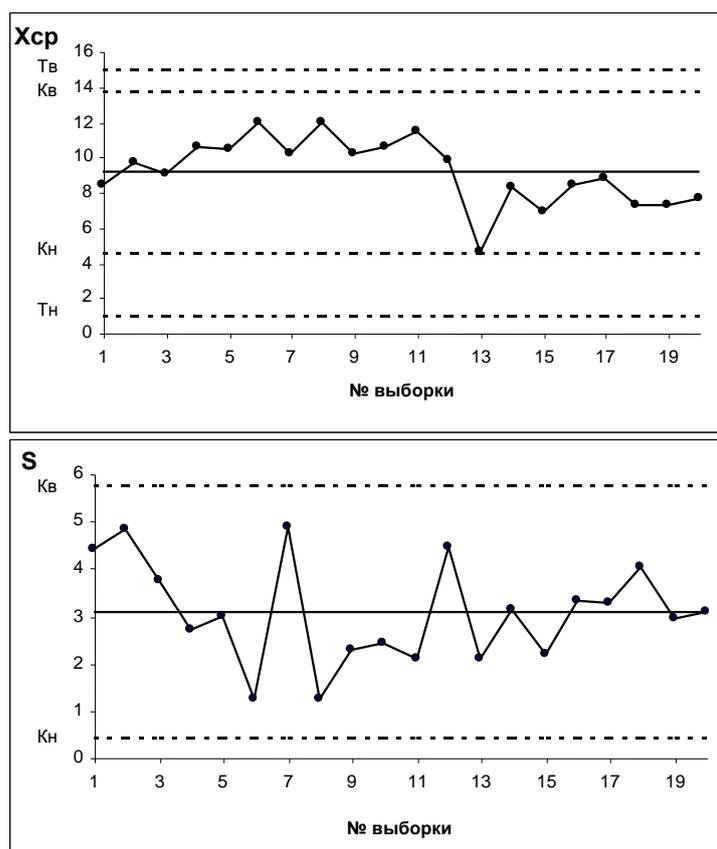


Рисунок 5.11 – $\bar{x} - s$ – карта

Анализ контрольной карты на рисунке 5.11 показывает, что рассеяние диаметра болта приемлемо, и по рассеянию процесс стабилен (оборудование настроено достаточно точно), поскольку на *s*-карте нет показаний разлаженности процесса. Однако на \bar{x} – карте

имеются серии из девяти точек (с четвёртой по двенадцатую) и из восьми точек (с тринадцатой по двадцатую), расположенных по одну сторону от средней линии. Это указывает на нестабильность процесса. Видимо, в течение процесса, при переходе от двенадцатой к тринадцатой точке изменилось математическое ожидание диаметра. Следует постараться выяснить причину этой нестабильности и провести управляющее воздействие на процесс. После стабилизации контрольную карту следует построить заново.

Порядок выполнения работы

Получив у преподавателя задание для выполнения лабораторной работы, студент изучает теоретические сведения. Далее строит контрольную карту, гистограмму и рассчитывает индексы производительности и пригодности процесса на компьютере. По описанной методике проводит анализ качества технологического процесса.

Содержание отчета

Отчёт по лабораторной работе должны содержать следующие пункты:

- название лабораторной работы;
- цель работы;
- индивидуальное задание для выполнения лабораторной работы;
- краткие теоретические сведения;
- краткое описание хода выполнения работы;
- результаты выполнения работы: контрольные карты с обозначенными зонами А, В и С; таблицы результатов тестов; таблица индексов производительности и пригодности, гистограмма с указанием границ поля допуска и границ регулирования процесса.
- подробные выводы по работе (о текущем состоянии и качестве технологического процесса).

Контрольные вопросы

5. Что такое контрольная карта и для чего она применяется?
6. Назовите типы контрольных карт.
7. Классификация контрольных карт по чувствительности к разладке технологического процесса.

8. Какие типы контрольных карт применяются при использовании количественных данных?
9. Проведите сравнение контрольных карт по чувствительности к разладке технологического процесса.
10. Каково значение контрольных карт в статистическом контроле качества?
11. Совместно с какими статистическими методами могут применяться контрольные карты?
12. Что означает термин «количественные признаки»?
13. Какие бывают виды контрольных карт по количественному признаку?
14. Какие значения необходимо рассчитать для построения контрольных карт по количественному признаку?
15. Какие признаки статистической неуправляемости можно обнаружить на контрольной карте?
16. Что характеризует индекс воспроизводимости процесса?
17. В чем особенности индекса настройки процесса?
18. Какие еще индексы и коэффициенты используются для анализа качества технологического процесса по контрольным картам, и в чем их особенности?

Лабораторная работа №6

Расчет измерителей качества

Цель работы: изучить получить основные сведения об измерителях качества и получить практические навыки их расчета в среде *LibreOffice*.

Задания для самостоятельного выполнения

Задание 1. Рассчитать уровень DPMO и PPM отдела работы с поставщиками, если известна следующая информация о поставках за последний отчетный период.

Задание 2. Оценить качество процесса (по данным полученным от преподавателя по каждому варианту) рассчитав измерители качества, если на выходном контроле после каждой операции было проконтролировано по 100 деталей. По каждой контролируемой характеристике показано количество несоответствий.

Задание 3. Определить вероятность прохождения каждой стадии процесса без единого дефекта (по данным полученным от преподавателя по каждому варианту), если на выходном контроле после каждой операции было проконтролировано по 200 деталей. По каждой контролируемой характеристике показано количество несоответствий.

Задание 4. Определить вероятность прохождения всех стадий процесса без единого дефекта (по данным полученным от преподавателя по каждому варианту), если на выходном контроле после каждой операции было проконтролировано по 100 деталей. По каждой контролируемой характеристике показано количество несоответствий.

Краткие теоретические сведения

Для измерения качества процесса следует использовать измерители качества и целевые показатели, представленные в Приложении Ж. Данные измерители качества могут использоваться в случаях, представленных в таблице 6.1.

Рассмотрим показатели качества и их связь с издержками на качество (см. таблицы 6.2 и 6.3).

Измерители качества, эффективно используемые в методологии Six Sigma – это DPU, DPO, DPMO и DPPM.

Таблица 6.1

Использование измерителей качества

Качество		
процесса в целом	технологической стадии	контролируемой характеристики
DPMO, sigma, $A_{СКВ}$	$A_{пр}$, D, A_1 , R, PPM, DPMO, sigma	C_p , C_{pk} , D, A_1 , R, PPM

Таблица 6.2

Корреляция с издержками измерителей качества

Корреляция с издержками	
сильная	слабая
D	A_1
DPMO	R
$A_{пр}$	PPM
$A_{СКВ}$	C_p
sigma	C_{pk}

Таблица 6.3

Издержки, связанные с качеством

Уровень сигма	DPMO	Издержки, связанные с качеством
2	308 537	Не рассчитывается
3	66 807	25-40% от выручки
4	6 210 (среднеотраслевой показатель)	15-20% от выручки
5	233	5-15% от выручки
6	3,4 (мировой класс)	Менее 1% от выручки
Увеличение качества процесса на 1 сигму обеспечивает рост чистого дохода на 10%		

DPU (Defect Per Unit) – показатель количества дефектов, приходящихся на единицу (продукции, услуги), рассчитывают по формуле:

$$DPU = \frac{\text{количество найденных дефектов}}{\text{количество проверенных единиц}} \quad (6.1)$$

Пример 1. Доставлено 100 коробок с пиццей, из которых в согласованные сроки – только 60. Остальные 40 – с задержкой, а потому – для заказчика считаются дефектом:

$$DPU = 40 / 100 = 0,4$$

Вместо коробок с пиццей могут быть:

- единицы изготовленной продукции, часть из которой содержит дефекты
- пакеты документов, оформленных без ошибок – дефектов
- какие-либо услуги, оказанные с огрехами.

DPO (Defect Per unit Opportunity) – показатель количества дефектов на одну возможность, можно рассчитать следующим образом:

$$DPU = \frac{DPU}{\text{количество возможностей на единицу}} \quad (6.2)$$

Пример 2. Количество дефектов на единицу продукции составляет 0,8. В то же время, количество ошибок, которое допускает отдел финансов при обработке документов, составляет 0,4. Допустим, предприятие выпускает корпуса автомобилей, состоящие из 50 деталей, которые крепятся между собой с помощью 1500 контактных точек (сваров). Документация считается неверно обработанной всякий раз, когда реквизиты предприятия, суммы либо значения любого другого поля заполнены неверно. Добавим грамматические ошибки и получим около 500 возможностей ошибки. Теперь рассчитаем показатели DPO для производственного процесса – 1 и процесса обработки документации – 2:

$$DPO_1 = 0,8 / (50 + 1500) = 0,000516$$

$$DPO_2 = 0,4 / 500 = 0,000800$$

Показатель DPO позволяет нам сравнить такие разные процессы как производство корпуса автомобиля и оформление финансовой документации, что дает возможность выделить участок для проведения проекта.

DPMO (Defect per Million Opportunity) – количество дефектов на миллион возможностей рассчитывают по формуле:

$$DPMO = DPO \times 1\,000\,000 \quad (6.3)$$

Как следует из формулы, показатель DPMO просто позволяет выразить определенную величину в более удобной форме, чем DPO. Для примера, рассмотренного выше, DPMO будет равно соответственно:

$$DPMO_1 = 0,000516 \times 1\,000\,000 = 516$$

$$DPMO_2 = 0,000800 \times 1\,000\,000 = 800$$

Пример 3. Проконтролировано 100 изделий, в результате выявлено 3 несоответствующих изделия по 1 несоответствию в каждом, при этом контролировалось 5 характеристик.

$$DPMO = \frac{3}{100 \cdot 5} 1000000 = 6000$$

$$PPM = \frac{3}{100} 1000000 = 30\,000$$

Всегда $DPMO < PPM$, кроме случая, когда выявлены несоответствия по каждому контролируемому параметру. В этом случае $DPMO = PPM$.

Рассчитав показатель DPMO можно:

1. Рассчитать σ – уровень процесса
2. Сравнить «яблоки с грушами»

Для того чтобы рассчитать сигма уровень процесса проще всего использовать таблицу перерасчета сигма уровней в количество дефектов на миллион возможностей (см. Приложение И), а установить их корреляцию с издержками, связанными с качеством можно по таблице 6.3. Для рассмотренного примера: процессу сварки корпуса соответствует уровень около 4,8, а процессу обработки документов – 4,65.

$A_{пр}$ характеризует этап технологического процесса. Так $A_{пр1} \times A_{пр2}$ характеризует вероятность прохождения 2-х стадий процесса без единого дефектов.

А $A_{скв}$ вероятность прохождения всех стадий процесса без единого дефекта. При этом данные показатели сильно связаны с издержками.

Пример 4. Рассмотрим одну из стадий обобщенного процесса представленную на рисунке 6.1.

а) При контроле по двум параметрам:

$$A_{пр2} = 1 - \frac{d_{высота} + d_{центрированмасс}}{N}$$

б) При контроле по одному параметру: $A_{првыс.} = 1 - \frac{d_{выс}}{N}$

$$A_{прцентриров.масс} = 1 - \frac{d_{центриров.масс}}{N}$$



Рисунок 6.1 – Одна из стадий обобщенного процесса

И тогда $A_{пр}$ для второй стадии можно записать в виде

$$A_{пр2} = A_{првыс} \times A_{прцентр.масс}$$

Пример 5. Фактический сквозной уровень процесса $A_{скв}^{факт} = 0,68$, а целевой сквозной уровень соответствия $A_{скв}^{цел} = 0,75$, то $A_{норм} = \sqrt[4]{0,75} = 0,87$, следовательно, если на каждой стадии будет уровень соответствия равный 0,87, то уровень соответствия всего процесса $A_{скв}^{цел} = 0,75$.

Порядок выполнения работы

Получив у преподавателя исходные данные для выполнения лабораторной работы, студент изучает теоретические сведения. Далее выполняет на компьютере расчет измерителей качества по формулам в Приложении Ж в программе *LibreOffice*.

Содержание отчета

Отчет по лабораторной работе должен содержать следующие пункты:

- название лабораторной работы;
- цель работы;
- краткие теоретические сведения;
- краткое описание хода выполнения работы;
- индивидуальные задания для выполнения лабораторной работы;
- результаты выполнения работы: таблицы с исходными данными, таблица расчета измерителей качества.
- выводы.

Контрольные вопросы

1. В каких единицах измеряется уровень несоответствий?
2. С помощью какого показатель можно определить вероятность прохождения всех стадий процесса без единого дефекта?
3. С помощью какого показатель можно определить вероятность прохождения каждой стадии процесса без единого дефекта?
4. Охарактеризуйте уровень соответствия при приемке с первого предъявления.
5. Что показывает доля несоответствующей продукции?
6. В чем заключается взаимосвязь уровня «сигма» и количества несоответствий на 1 млн. возможностей появления этих несоответствий?
7. В чем отличие DPMO и PPM?

Лабораторная работа №7

Методы анализа влияния факторов

Цель работы: получить основные сведения о методах анализа влияния факторов и получить практические навыки их использования, а также навыки расчетов и построения графиков при проведении анализа в среде *LibreOffice*.

Задания для самостоятельного выполнения:

Задание 1. Выяснить, есть ли существенные различия в работе двух смен, на основании данных выданных преподавателем.

Какой графический метод следует выбрать для представления результатов анализа?

Задание 2. Выяснить, как влияет замена шлифовального круга на результаты измерений на основании данных выданных преподавателем. Какой графический метод следует выбрать для представления результатов анализа?

Задание 3. Выяснить, как влияет замена шлифовального круга на долю брака в первую и вторую смены на основании данных выданных преподавателем. Какой графический метод следует выбрать для представления результатов анализа?

Задание 4. Необходимо оценить работу двух лаборантов, на основании данных выданных преподавателем. Какой графический метод следует выбрать для представления результатов анализа?

Задание 5. На основании данных выданных преподавателем проанализировать данные экспертов и проверить согласованность их мнений.

Задание 6. Выяснить, оказало ли влияние изменение технологии обработки поверхности изделия на контролируемую характеристику, на основании данных выданных преподавателем. Какой графический метод следует выбрать для представления результатов анализа?

Задание 7. Оценить использование трех станков разной модификации на точность изготовления детали на основании данных выданных преподавателем.

Задание 8. Выяснить, как влияет замена шлифовального круга на количество брака (шт.) полученного в первую и вторую смены на основании данных выданных преподавателем. Какой графический метод следует выбрать для представления результатов

анализа?

Задание 9. Была разработана новая методика. Для оценки ее эффективности были проведены контрольные опыты (данные выдаются преподавателем). Можно ли рекомендовать применять новую методику вместо общепринятой? Какой графический метод следует выбрать для представления результатов анализа?

Задание 10. Два лаборанта проводили измерения. На основании данных выданных преподавателем определить можно ли объединить результаты лаборантов в одну выборку? Какой графический метод следует выбрать для оценки результатов работы лаборантов?

Краткие теоретические сведения

Существует довольно много подходов, позволяющих определить, оказывает ли некоторый фактор влияние на качество процесса. Например, мы говорим, что уровень квалификации персонала оказывает влияние на качество выпускаемой продукции. Так ли это? Как доказать наличие данного влияния? И если влияние есть, то насколько сильное? Или, как доказать, что качество сырья оказывает влияние на качество продукции? Конечно, есть мнения экспертов, но хотелось бы получить более объективную доказательную базу для выводов. Для решения подобных вопросов активно используют статистические методы.

Методы, позволяющие подтвердить влияние фактора можно классифицировать следующим образом (рисунок 7.1):



Рисунок 7.1 – Классификация методов анализа влияния факторов

С помощью аналитических методов анализа можно решить большинство практических задач по анализу факторных влияний на

качество продукции. Все методы объединены одной целью – выяснить оказывает ли влияние некоторый фактор на качество продукции или процесса. Обратите внимание, что методы не отвечают на вопрос о том, какое влияние оказывает фактор. Чтобы оценить силу влияния фактора можно использовать графический инструментарий.

Чтобы выбрать адекватный метод анализа в любой практической ситуации следует использовать схему выбора метода анализа влияния факторов (рисунок 7.2).



Рисунок 7.2 – Схема выбора метода анализа влияния фактора

Использование графических методов для представления результатов анализа, представлено в таблице 7.1.

Таблица 7.1

Использование графических методов для представления результатов анализа

Метод	Вид графика
T – критерий Стьюдента	«Ящик с усами» (box plot)
Корреляция Спирмена	Диаграмма рассеяния
Критерий Краскела-Уоллиса	«Ящик с усами» (box plot)
Критерий Хи-квадрат	Круговая диаграмма

Критерий «хи-квадрат» для анализа таблиц сопряженности

Исходные данные для анализа:

- 1) Любая ячейка таблицы сопряженности должна содержать не менее 5 наблюдений;
- 2) Количество уровней номинальной шкалы не должно превышать 4. В противном случае результаты анализа будут утрачивать информативность.

Критерий хи-квадрат определяется по формуле:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}, \quad (7.1)$$

где f_o – наблюдаемая частота в каждой ячейке; f_e – ожидаемые частоты в каждой ячейке.

Ожидаемые частоты определяются исходя из данных таблицы сопряженности наблюдаемых частот по формуле:

$$f_e = \frac{n_r \cdot n_c}{n}, \quad (7.2)$$

где n_r – накопительная частота по строке; n_c – накопительная частота по столбцу; n – общий размер выборки.

Корреляция Спирмена и Кендалла

Для ранговых данных используются другие коэффициенты корреляции, такие как коэффициенты Спирмена и Кендалла.

Коэффициент ранговой корреляции Спирмена определяется по формуле:

$$K_C = 1 - 6 \sum_{i=1}^n \frac{d_i^2}{n(n^2 - 1)} \quad (7.3)$$

где n – количество наблюдений; d_i – разность ранговых показателей в i -том наблюдении.

В отличие от коэффициента Пирсона данный коэффициент не является мерой линейной связи. Коэффициент K_C лежит в промежутке от -1 до 1. При совпадении рангов $K_C = 1$, при противоположных рангах $K_C = -1$.

Значимость коэффициента Спирмена может быть проверена по t-критерию

$$t = K_C \sqrt{\frac{n-2}{1-K_C^2}}, \quad (7.4)$$

имеющему распределение Стьюдента с числом степеней свободы $n - 2$.

Коэффициент ранговой корреляции Кендалла определяется по формуле:

$$\tau = 1 - \frac{4Q}{n(n-1)}, \quad (7.5)$$

где Q – минимальное число обменов соседних элементов одной из ранжировок для ее приведения (совпадения) с другой.

Как и для коэффициента Спирмена при совпадении рангов $\tau = 1$, при противоположных рангах $\tau = -1$.

Статистика для проверки значимости этого коэффициента

$$Z = \sqrt{\frac{9n(n-1)}{2(2n+5)}} \tau \quad (7.6)$$

имеет нормальное распределение $N(0, 1)$.

Коэффициент ранговой корреляции Кендалла имеет некоторые преимущества перед коэффициентом Спирмена – в частности, он может использоваться и для многофакторного анализа. При достаточно большом числе объектов ($n \geq 10$) между значениями ранговых коэффициентов существует простая связь $K_C = 1,5\tau$.

Т-критерий для независимых выборок

Исходные данные для анализа:

- 1) Номинальная шкала должна иметь ровно 2 уровня;
- 2) Данные внутри сравниваемых подгрупп должны быть нормально распределены;

3) Каждый уровень номинальной шкалы должен иметь хотя бы 30 значений.

Анализ проводится по T-критерию Стьюдента:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_{\bar{x}_1}^2}{n_1} + \frac{s_{\bar{x}_2}^2}{n_2}}}, \quad (7.7)$$

где n_1, n_2 – размеры сравниваемых выборок;

\bar{x}_1, \bar{x}_2 – среднее значение по выборкам;

$s_{\bar{x}_1}, s_{\bar{x}_2}$ – выборочные стандартные ошибки и определяются по формуле:

$$s_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (7.8)$$

Расчетное значение T-критерия Стьюдента сравниваем с табличным.

Определяем $T_{\text{табл}}$ по значению числа степеней свободы
число степеней свободы = $n_1 + n_2 - 2$. (7.9)

Если расчетное значение больше табличного, то фактор оказывает существенное влияние на зависимую величину.

Дисперсионный анализ

Во многих практических ситуациях представляет интерес влияние того или иного качественного фактора на рассматриваемый показатель. Влияет ли квалификация наладчиков на качество обработки поверхности? Влияет ли метод обработки на точность изготовления детали? Зависит ли прочность болта из стекловолокнита от температуры при прессовании? Ответ на эти и аналогичные вопросы дается методами однофакторного дисперсионного анализа.

Пусть, например, оценка качества поверхности детали проводится с помощью k приборов и необходимо исследовать, влияет ли фактор «прибор» на результат измерений. Если приборов два, то проверка гипотезы о равенстве средних показаний приборов проводится рассмотренными выше методами. Если же приборов больше двух, используются методы дисперсионного анализа. Проверяется нулевая гипотеза $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ об отсутствии влияния на результативный признак X (результат измерений)

фактора А («прибор»), имеющего k уровней $A_j, j = 1, 2, \dots, k$.

Основная идея дисперсионного анализа состоит в том, чтобы сопоставить дисперсию за счет воздействия фактора А с дисперсией, обусловленной случайными причинами. Если различие между ними несущественно, то влияние фактора А на признак Х незначительно. Если же различие между факторной и остаточной дисперсиями значимо, то это говорит о влиянии фактора А на рассматриваемый признак Х.

Предполагается, что случайная величина Х имеет нормальное распределение с математическим ожиданием μ_j , зависящим от уровня фактора A_j , и постоянной дисперсией σ^2 . В качестве исходных данных используются выборочные значения величины Х, полученные для каждого уровня фактора А; число элементов выборки на каждом уровне равно n , тогда общее число наблюдений nk , обозначим через x_{ij} результат i -го наблюдения ($i = 1, 2, \dots, n$) за j -м фактором.

Выборочное среднее, соответствующее j -му уровню фактора А (групповое среднее), вычисляется по формуле

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad (7.10)$$

а общее среднее

$$\bar{x} = \frac{1}{nk} \sum_{j=1}^k \sum_{i=1}^n x_{ij} = \frac{1}{k} \sum_{j=1}^k \bar{x}_j \quad (7.11)$$

Общая сумма квадратов — это сумма квадратов отклонений наблюдаемых значений x_{ij} от общего среднего:

$$Q = \sum_{j=1}^k \sum_{i=1}^n (x_{ij} - \bar{x})^2 = \sum_{j=1}^k \sum_{i=1}^n x_{ij}^2 - nk\bar{x}^2 \quad (7.12)$$

Факторная сумма квадратов (обусловленная влиянием фактора А) — это сумма квадратов отклонений групповых средних от общей средней:

$$Q_A = n \sum_{j=1}^k (\bar{x}_j - \bar{x})^2 = n \sum_{j=1}^k \bar{x}_j^2 - nk\bar{x}^2 \quad (7.13)$$

Остаточная сумма квадратов характеризует рассеяние внутри группы:

$$Q_e = \sum_{j=1}^k \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad (7.14)$$

На практике эта сумма определяется из основного тождества дисперсионного анализа, в соответствии с которым

$$Q = Q_A + Q_e \quad (7.15)$$

Соответствующие числа степеней свободы

$$v = nk - 1; v_A = k - 1; v_e = k(n - 1) \quad (7.16)$$

а дисперсии

$$s^2 = Q/v; s_A^2 = Q_A/v_A; s_e^2 = Q_e/v_e \quad (7.17)$$

Если нулевая гипотеза о равенстве средних справедлива, то эти дисперсии являются несмещенными оценками дисперсий генеральной совокупности. Значительное превышение дисперсии s_A^2 над дисперсией s_e^2 можно объяснить различием средних в группах. Поэтому для проверки нулевой гипотезы используется статистика

$$F = \frac{s_A^2}{s_e^2} = \frac{Q_A/(k-1)}{Q_e/k(n-1)}, \quad (7.18)$$

которая имеет распределение Фишера с числами степеней свободы $(k - 1)$ и $k(n - 1)$. Нулевая гипотеза не противоречит результатам наблюдений на заданном уровне значимости α , если

$$F < F_{1-\alpha}(k-1, k(n-1)). \quad (7.19)$$

В этом случае считается, что фактор А не оказывает существенного влияния на показатель Х.

Результаты расчета сводятся в таблицу 7.2.

Таблица 7.2

Результаты проведения дисперсионного анализа

Источник дисперсии	Сумма квадратов	Число степеней	Дисперсия	Статистика Фишера
Фактор А	Q_A	v_A	s_A^2	F
Остаток	Q_e	v_e	s_e^2	
Общая	Q	v	s^2	

U-критерий Манна-Уитни

U-критерий Манна-Уитни (англ. Mann-Whitney U test) — непараметрический статистический критерий, используемый для оценки различий между двумя выборками по уровню какого-либо

признака, измеренного количественно. Позволяет выявлять различия в значении параметра между малыми выборками.

Этот метод определяет, достаточно ли мала зона перекрещивающихся значений между двумя рядами (ранжированным рядом значений параметра в первой выборке и таким же во второй выборке). Чем меньше значение критерия, тем вероятнее, что различия между значениями параметра в выборках достоверны.

Ограничения применимости критерия

1. 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.

2. В каждой выборке должно быть не более 60 значений параметра, но уже при выборках в 20 и более единиц ранжирование становится довольно трудоемким.

Использование критерия

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

1. Составить единый ранжированный ряд из обоих сопоставляемых выборок, расставив их элементы по степени нарастания признака и приписав меньшему значению меньший ранг. Общее количество рангов получится равным: $N = n_1 + n_2$, где n_1 — количество единиц в первой выборке, а n_2 — 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 (T_x), соответствующую выборке с n_x единиц.

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

$$U = n_1 \cdot n_2 + \frac{n_x \cdot (n_x + 1)}{2} - T_x.$$

4. По таблице определить критические значения критерия для данных n_1 и n_2 . Если полученное значение U меньше табличного или равно ему для избранного уровня статистической

значимости, 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 .

Критерий Краскела-Уоллиса

Если мы не можем сказать что-либо определенное об альтернативах к H_0 , можно воспользоваться для ее проверки свободным от распределения критерием Краскела–Уоллиса. Для этого заменим наблюдения x_{ij} их рангами r_{ij} , упорядочивая всю совокупность $\|x_{ij}\|$ в порядке возрастания (для определенности). Затем для каждой обработки j (т.е. для каждого столбца исходной таблицы) надо вычислить

$$R_j = \sum_{i=1}^{n_j} r_{ij}$$

где R_j — это средний ранг, рассчитанный по столбцу. Если между столбцами нет систематических различий, средние ранги R_j , $j = 1, \dots, k$ не должны значительно отличаться от среднего ранга, рассчитанного по всей совокупности $\|r_{ij}\|$. Ясно, что последний равен $(N + 1)/2$. Поэтому величины

$$\left(R_1 - \frac{N+1}{2}\right)^2, \dots, \left(R_k - \frac{N+1}{2}\right)^2$$

при H_0 в совокупности должны быть небольшими. Составляя общую характеристику, разумно учесть различия в числе наблюдений для разных обработок и взять в качестве меры отступления от чистой случайности величину

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)$$

Эта величина называется статистикой Краскела–Уоллеса. Множитель $12/[N(N + 1)]$ нужен для стабилизации ее распределения при большом числе наблюдений.

Примеры выполнения заданий

Пример 1: По данным количества продукции 1 и 2 сорта в разных цехах составляем таблицу сопряженности наблюдаемых частот (таблица 7.3)

Таблица 7.3

Таблица сопряженности наблюдаемых частот

Сорт	Номер цеха		Сумма
	Цех №1	Цех №2	
Сорт 1	34	48	82
Сорт 2	5	23	28
Сумма	39	71	110

По формуле (7.2) рассчитываем ожидаемые частоты и составляем таблицу сопряженности ожидаемых теоретических частот (таблица 7.4).

Таблица 7.4

Таблица сопряженности ожидаемых теоретических частот

Сорт	Номер цеха		Сумма
	Цех №1	Цех №2	
Сорт 1	29,1	52,9	82
Сорт 2	9,9	18,1	28
Сумма	39	71	110

$$f_{e11} = \frac{82 \cdot 39}{110} = 29,1$$

$$f_{e12} = \frac{28 \cdot 39}{110} = 9,9$$

$$f_{e21} = \frac{82 \cdot 71}{110} = 52,9$$

$$f_{e22} = \frac{28 \cdot 71}{110} = 18,1$$

И на основании данных таблиц 7.3 – 7.4 рассчитываем критерий хи-квадрат по формуле (7.1):

$$\chi_{расч}^2 = \frac{(34 - 29,1)^2}{29,1} + \frac{(48 - 52,9)^2}{52,9} + \frac{(23 - 18,1)^2}{18,1} + \frac{(5 - 9,9)^2}{9,9} = 5,22$$

Расчетное значение критерия хи-квадрат сравниваем с табличным.

Определяем $\chi_{табл}^2$ по значению числа степеней свободы **число степеней свободы = (число строк – 1)(число столбцов – 1)**, в нашем случае число степеней свободы равно 1, и при уровне значимости 5%, $\chi_{табл}^2 = 3,841$.

Если расчетное значение больше табличного, то фактор оказывает существенное влияние на зависимую величину с вероятностью 95%.

Как и в нашем примере, $\chi_{расч}^2 > \chi_{табл}^2$, следовательно, цех №2

работал хуже, так как выпустил больше продукции 2 сорта.

Для графического представления данных используется круговая диаграмма (рисунок 7.3).

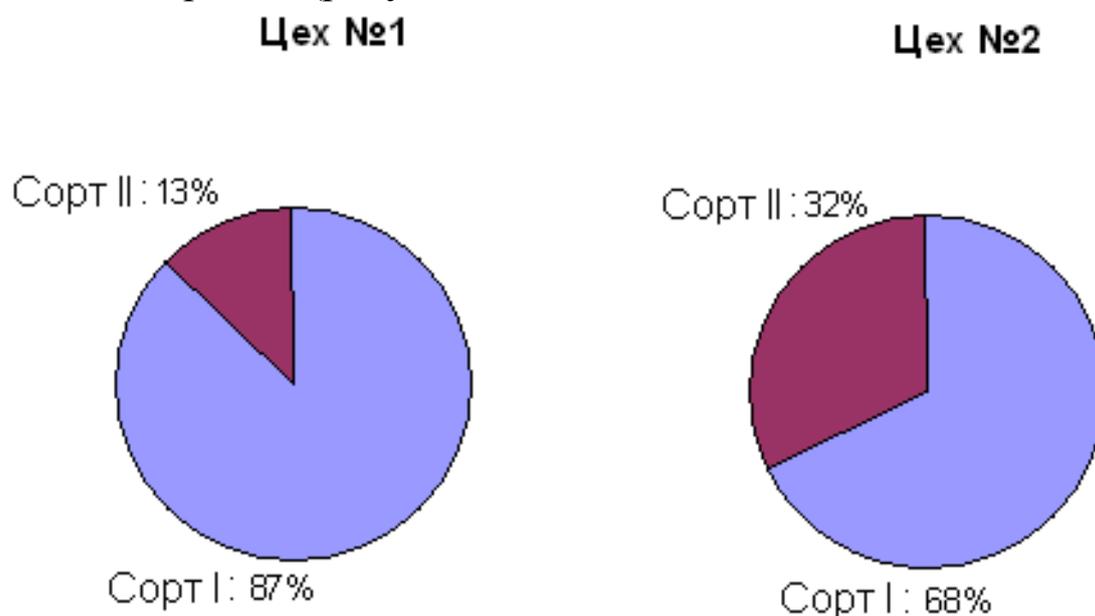


Рисунок 7.3 – Круговая диаграмма

Пример 2. Два эксперта проводят органолептический анализ жемчуга: ранжируют по убыванию качества 10 жемчужин. Результаты представлены в первых трех строках таблицы 7.5.

Таблица 7.5

№ п/п	1	2	3	4	5	6	7	8	9	10
Эксперт 1	2	1	3	4	6	5	8	7	10	9
Эксперт 2	3	2	1	4	6	7	5	9	10	8
d_i	-1	-1	2	0	0	-2	3	-2	0	1
d_i^2	1	1	4	0	0	4	9	4	0	1

Проверить согласованность мнений экспертов по коэффициенту ранговой корреляции Спирмена.

Решение:

Найдем входящие в формулу (7.3) величины d_i - разность между рангами, присвоенные экспертами i -му объекту, а также квадраты этих величин; результаты вычислений приведены в нижних двух строках таблицы. Тогда

$$K_C = 1 - 6 \cdot 24 / 990 = 0,85$$

Для проверки значимости найдем значение статистики Стьюдента по формуле (7.4):

$$t_{расч} = 0,85 \sqrt{\frac{10-2}{1-0,85^2}} = 4,56$$

Критическое значение при уровне значимости 0,05 по справочной таблице $t_{кр} = 1,86$. Так как $t_{расч} > t_{кр}$ - корреляция значима. Учитывая достаточно высокое значение коэффициента корреляции и его значимость, можно считать, что степень близости ранжировок экспертов высокая.

Пример 3. Необходимо оценить влияние технологии чистовой обработки (три вида технологий) на точность изготовления детали. При каждом виде технологии проводится по четыре замера отклонения размера детали от номинала в мкм (см. таблицу 7.6).

Таблица 7.6

Номер замера	Вид технологии		
	1	2	3
1	1	2	3
2	2	1	2
3	2	3	2
4	1	2	3

Решение:

Имеем $n = 4$, $k = 3$.

Групповые средние

$$\bar{x}_1 = (1+2+2+1)/4 = 1,5; \quad \bar{x}_2 = (2+1+3+2)/4 = 2; \quad \bar{x}_3 = (3+2+2+3)/4 = 2,5,$$

а

общая средняя $\bar{x} = (1,5+2+2,5)/3=2$.

Общая сумма квадратов

$$Q = 1^2 + 2^2 + 2^2 + 1^2 + 2^2 + 1^2 + 3^2 + 2^2 + 3^2 + 2^2 + 2^2 + 3^2 - 4 \cdot 3 \cdot 2^2 = 6,$$

факторная сумма квадратов

$$Q_A = 4 (1,5^2 + 2^2 + 2,5^2) - 4 \cdot 3 \cdot 2^2 = 2,$$

остаточная сумма квадратов

$$Q_e = Q - Q_A = 4.$$

Результаты расчета сводим в таблицу 7.7.

Из справочной таблицы находим квантиль распределения Фишера

$$F_{1-\alpha}(k-1, k(n-1)) = F_{0,95}(2,9) = 4,26.$$

Таблица 7.7

Источник дисперсии	Сумма квадратов	Число степеней	Дисперсия	Статистика Фишера
Фактор А	2	2	1	2,25
Остаток	4	9	0,44	
Общая	6	11	0,55	

Так как выборочное значение статистики оказалось меньше критического (рисунок 7.4), нулевая гипотеза принимается: в данном случае влияние технологии изготовления на точность детали несущественно.

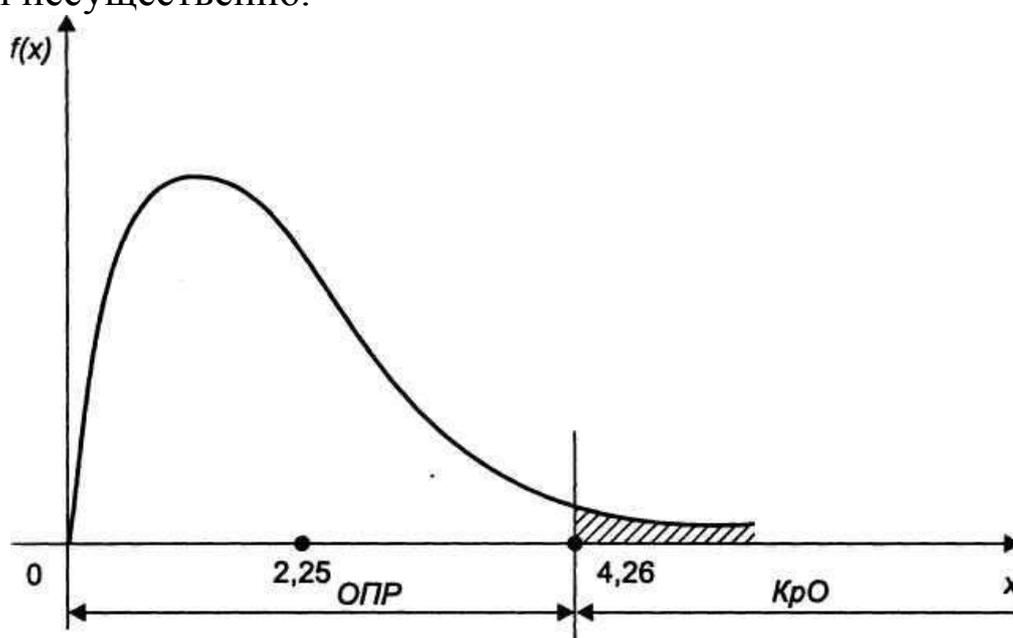


Рисунок 7.4 – Положение критической области при проведении дисперсионного анализа

Подобным образом в двухфакторном дисперсионном анализе оценивается влияние двух факторов А и В, а также их взаимодействия АВ на результативный признак Х: проверяются три соответствующие нулевые гипотезы. В трехфакторном анализе по аналогии исследуется влияние на признак Х трех факторов А, В и С, их парных взаимодействий АВ, ВС и АС, а также общего взаимодействия АВС.

Пример 4. Необходимо установить оказало ли влияние замена шлифовального круга на контролируемую характеристику, если известно, что «до замены» была взята выборка объемом 60 единиц, среднее арифметическое которой равно 24,99960 мм, стандартное

отклонение – 0,000924 мм; «после замены» была взята выборка объемом 90 единиц при этом среднее арифметическое стало равно 24,99952 мм, а дисперсия – 0,000986. Какой графический метод следует выбрать для представления результатов анализа?

Решение:

По полученным данным выдвигаем гипотезу о том, что замена шлифовального круга повлияла на процесс шлифования колец подшипников. Данную гипотезу проверим, рассчитав Т-критерий Стьюдента.

Выборочные стандартные ошибки определяем по формуле (7.8):

$$s_{\bar{x}_1}^2 = \frac{\sigma^2}{n} = \frac{0,000924^2}{60} = 1,42 \cdot 10^{-8}$$

$$s_{\bar{x}_2}^2 = \frac{\sigma^2}{n} = \frac{0,000986^2}{90} = 1,08 \cdot 10^{-8}$$

Подставим полученные характеристики в формулу (7.7) для расчета критерия Стьюдента:

$$T_{расч} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_{\bar{x}_1}^2}{n_1} + \frac{s_{\bar{x}_2}^2}{n_2}}} = \frac{24,9996 - 24,99952}{\sqrt{\frac{1,42 \cdot 10^{-8}}{60} + \frac{1,08 \cdot 10^{-8}}{90}}} = \frac{0,8 \cdot 10^{-4}}{0,189 \cdot 10^{-4}} = 4,23$$

По справочной таблицы значений Т-критерию Стьюдента определяем табличное значение критерия при уровне значимости 0,05 и числе степеней свободы 148 по формуле (7.9), $T_{табл} = 1,9759$.

Поскольку $T_{расч} > T_{табл}$, то фактор не оказывает существенное влияние на зависимую величину, следовательно, замена шлифовального круга оказывает существенное влияние на замену шлифовального круга при шлифовании колец подшипников.

Для графического представления данных используем «ящик с усами» или диаграмму размаха (рисунок 7.5).

Порядок выполнения работы

Получив у преподавателя исходные данные для выполнения лабораторной работы, студент изучает теоретические сведения согласно пункту 3. Далее выполняет на компьютере задания на основании исходных данных выданных преподавателем в программе *LibreOffice*.

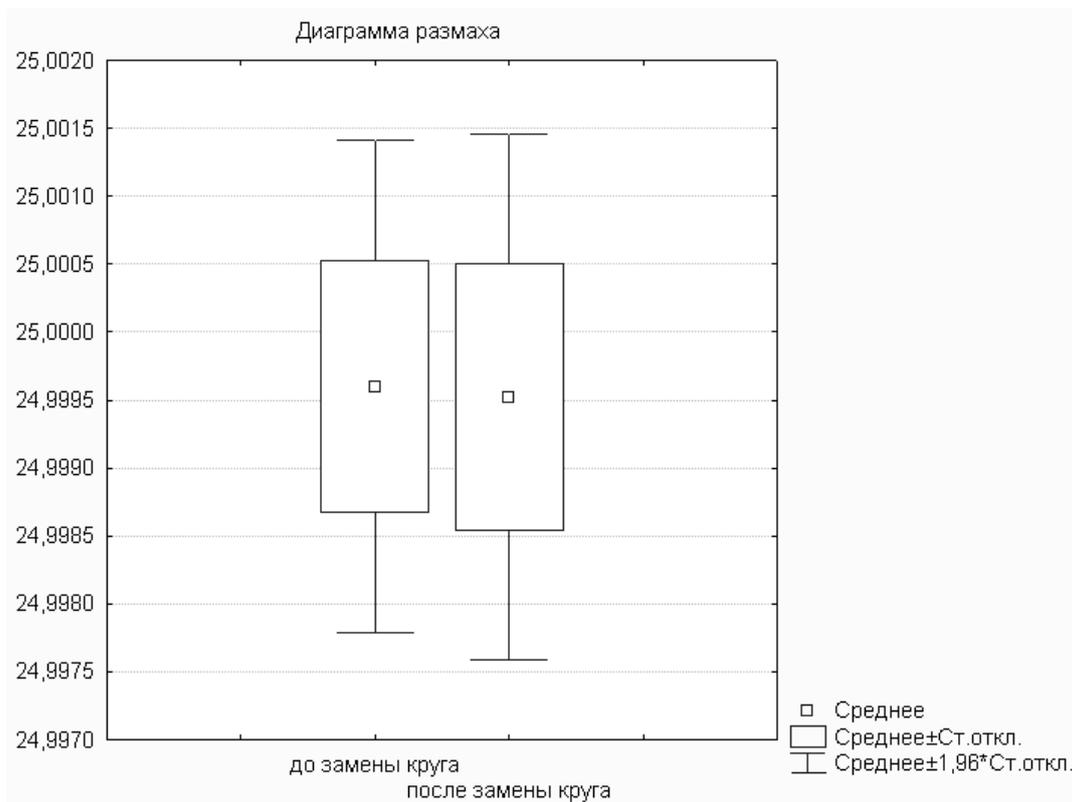


Рисунок 7.5 – «Ящик с усами» для оценки размаха диаметра колец подшипников до и после замены шлифовального круга

Содержание отчета

Отчет по лабораторной работе должен содержать следующие пункты:

- название лабораторной работы;
- цель работы;
- краткие теоретические сведения;
- краткое описание хода выполнения работы;
- индивидуальные задания для выполнения лабораторной работы;
- результаты выполнения работы: таблицы с исходными данными, графики, итоговая таблица расчета.
- выводы.

Контрольные вопросы

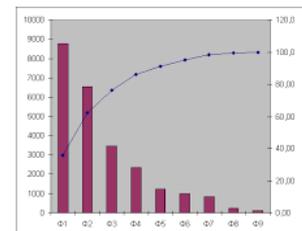
1. Перечислите основные этапы алгоритма проверки гипотез о параметрах распределения.
2. Какие основные распределения, используемые в статистических расчетах Вы знаете?

3. Как определяют квантили распределений?
4. При решении, каких задач используются методы дисперсионного анализа?
5. Что понимают под адекватностью регрессионной модели?
6. Что характеризует коэффициент детерминации?
7. Какова зависимость между значениями ранговых коэффициентов Кендела и Спирмена?
8. Для анализа влияния, каких факторов используется критерий Стьюдента?

Тестовые задания для самоконтроля

1. На рисунке представлен график

- а) Диаграммы Парето
- б) Дома качества
- в) Сравнения планируемого и фактического

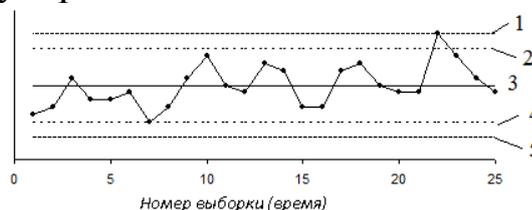


выполнения плана

г) Диаграммы Исикавы

2. Контрольные карты позволяют проводить анализ _____ процесса.

3. Установите последовательность расположения границ на контрольной карте Шухарта:



- а) верхняя граница регулирования
- б) верхняя граница поля допуска
- в) нижняя граница регулирования
- г) нижняя граница поля допуска
- д) центральная линия и/или номинальное значение

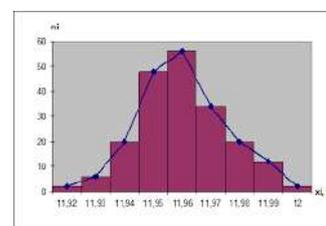
4. Установите соответствие:

- 1. Индекс пригодности
- 2. Индекс воспроизводимости
- 3. Наименьший индекс воспроизводимости
- 4. Верхний индекс пригодности
- 5. Коэффициент точности

- а) CP
- б) PP
- в) CR
- г) CPK
- д) PPKU

5. На рисунке изображено ...

- а) Гистограмма
- б) Диаграмма рассеяния
- в) Диаграмма разброса
- г) Диаграмма Парето



6. Измерители качества, эффективно используемые в методологии *Six Sigma* – это

- а) *DPU*
- б) *DPO*
- в) *DPMO*
- г) *DPPM*
- д) все перечисленные

7. Какой показатель характеризует вероятность прохождения j -ой стадии процесса без единого дефекта?

- а) Промежуточный уровень соответствия
- б) Сквозной уровень соответствия
- в) Нормализованный уровень соответствия
- г) DPMO
- д) PPM

8. Какой показатель характеризует вероятность прохождения всех стадий процесса без единого дефекта? Ответ: _____ .

9. Контроль – это процесс определения соответствия...

- а) значения параметра изделия установленным требованиям или нормам;
- б) значения параметра изделия техническим требованиям;
- в) значения параметра изделия требованиям, регламентированным ГОСТ Р;
- г) значения параметра изделия техническим условиям.

10. Контроль продукции или процесса во время выполнения или после завершения технологической операции, называется _____ контролем.

11. Эксперимент, в процессе которого исследуется стохастическая зависимость одной величины Y от нескольких других величин X_i , называется _____ .

12. Установите последовательность этапов проверки статистических гипотез:

- 1 – формулировка основной гипотезы
- 2 – расчёт статистики критерия
- 3 – задание уровня значимости
- 4 – построение критической области
- 5 – выводы об истинности гипотезы

13. Установите соответствие

1	Коэффициент регрессии	А	$R = \frac{\text{cov}(X, Y)}{\sigma_x \cdot \sigma_y}$
2	Линейный коэффициент корреляции	Б	$R_{y/x} = \frac{\sigma_y}{\sigma_x} r$
3	Выборочный линейный коэффициент корреляции	В	$r_{xy}^p = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}$

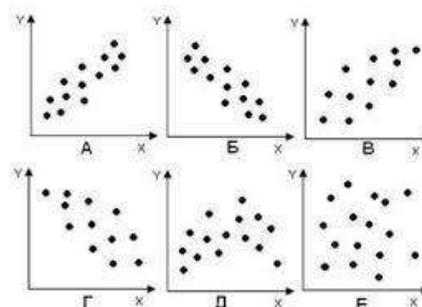
4	Коэффициента корреляции Пирсона	$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$
---	------------------------------------	--

14. Установите соответствие

Метод	Вид графика
T – критерий Стьюдента	«Ящик с усами» (box plot)
Корреляция Спирмена	«Ящик с усами» (box plot)
Критерий Краскела-Уоллиса	Круговая диаграмма
Критерий Хи-квадрат	Диаграмма рассеяния

15. Установите соответствие:

- 1 – сильная положительная корреляция
- 2 – слабая положительная корреляция
- 3 – сильная отрицательная корреляция
- 4 – слабая отрицательная корреляция
- 5 – отсутствие корреляции
- 6 – криволинейная корреляция



16. Чертеж, на котором статистическая информация изображается посредством геометрических фигур или символических знаков называется _____ .

17. Установите последовательность основных этапов статистического управления качеством:

- 1 – статистическое обследование
- 2 – наладка процесса
- 3 – статистическое управление

Библиографический список

1. Клячкин, В. Н. Статистические методы в управлении качеством: компьютерные технологии : учебное пособие / В. Н. Клячкин. - Москва : Финансы и статистика, 2009. - 304 с. : ил.

2. Статистические методы контроля и управления качеством: методические указания по выполнению практических работ и самостоятельной работы по дисциплине «Статистические методы контроля и управления качеством» для студентов направления подготовки 27.03.01 Стандартизация и метрология / Юго-Зап. гос. ун-т; сост.: С.В. Ходыревская. Курск, 2018. 144 с.: прилож. 13. Библиогр.: с. 119.

3. Инструменты контроля и управления качеством: методические указания по выполнению лабораторной и самостоятельной работы бакалавров и магистров, изучающих дисциплины «Управление качеством» и «Управление качеством продукции» / Юго-Зап. гос. ун-т; сост.: С.В. Ходыревская. Курск, 2018. 34 с. Библиогр.: с. 33.

4. Эванс, Д. Управление качеством [Электронный ресурс] : учебное пособие / Д. Эванс. - Москва : Юнити-Дана, 2015. - 671 с // Режим доступа - <http://biblioclub.ru>

5. Статистические методы повышения качества /Под ред. Хитоси Кумэ – М.: Финансы и статистика, 1990. 147 с.: ил.

6. «Семь инструментов качества» в японской экономике –М.: Изд-во стандартов, 1990. 89 с.: ил.

7. Исикава К. Японские методы управления качеством : Сокр. пер. с англ./Науч. ред. и авт. предисл. А.В. Гличев. - М: Экономика, 1988. - 215 с.

8. Умарова, Н. Н. Статистические методы в управлении качеством (использование программного продукта STATISTICA) [Электронный ресурс] : учебно-методическое пособие / Н. Н. Умарова, Р. Ф. Бакеева. - Казань : КГТУ, 2008. - 112 с. // Режим доступа - <http://biblioclub.ru>

9. Пономарев С. В. Управление качеством продукции. Введение в системы менеджмента качества: Учебное пособие / С. В. Пономарев, С. В. Мищенко, В. Я. Белобрагин. - М. : Стандарты и качество, 2004. - 248 с. (гриф УМО)

10. Анализ точности и стабильности процессов : учебное пособие / Ю. М. Быков [и др.]. - Старый Оскол : ТНТ, 2014. - 96 с.
11. Ефимов, В. В. Статистические методы в управлении качеством продукции : учебное пособие / В. В. Ефимов, Т. В. Барт. - М. : КноРус, 2006. - 240 с.
12. Жулинский, С. Ф. Статистические методы в современном менеджменте качества / С. Ф. Жулинский, Е. С. Новиков. - М. : Новое тысячелетие, 2001. - 206 с.
13. Дуброва, Т. А. Статистические методы прогнозирования : учебное пособие / Т. А. Дуброва. - М. : Юнити, 2003. - 206 с.
14. ГОСТ Р ИСО 7870-2-2015 Статистические методы. Контрольные карты. Часть 2. Контрольные карты Шухарта. Издание официальное. – М.: Стандартиформ, 2016. 42 с.
15. ГОСТ Р ИСО 22514-1-2015 Статистические методы. Управление процессами. Часть 1. Общие принципы. Издание официальное. – М.: Стандартиформ, 2016. 22 с.
16. Все о качестве. Отечественные разработки. Вып.6. «Статистическое управление технологическим процессом». –М.: НТК «Трек», 2002. 60 с.

Приложение Б

Контрольный листок видов дефектов

Контрольный листок		
Наименование изделия: _____		Дата: _____
Производственная операция: приемочный контроль		Участок: _____
Тип дефекта: царапина, пропуск операции, трещина, неправильная обработка		Фамилия контролера: _____
Общее число проконтролированных изделий: 1525		Номер партии: _____
Примечания: по всем проконтролированным изделиям		Номер заказа: _____
Тип дефекта	Результат контроля	Итоги по типам дефектов
Поверхностные царапины	### ### ### //	17
Трещины	### ### /	11
Пропуск операции	### ### ### ### ### /	26
Неправильное исполнение операции	///	3
Другие	###	5
	Итого	62
Общее число забракованных изделий	### ### ### ### ### ### ### ### //	42

Приложение Г

Контрольный листок причин дефектов

Оборудование	Рабочий	Пон.			Вт.			Ср.			Чт.			Пт.			Суб.		
		до перерыва	после перерыва	до перерыва	после перерыва	до перерыва	после перерыва	до перерыва	после перерыва	до перерыва	после перерыва	до перерыва	после перерыва	до перерыва	после перерыва	до перерыва	после перерыва		
Станок 1	A	oo x •	o x	ooo	o xx	ooo xxx	ooo xxx	oooo xxx	oooo x	oooo xx	o	oooo	oo	o	oo	oo	xx		
	B	o xx •	ooo xxx •	ooooo xx	ooo xx	ooooo xx	ooooo x	ooooo xx	ooooo x	ooooo xx	ooo	oo	ooooo	oo	ooooo	oo	xx		
Станок 2	C	oo x	o x	oo	•	ooooo	ooooo x	oo	o	oo	oo	oo	oo	oo	oo	oo	oo		
	D	oo x	o x	oo Δ	ooo •	ooo Δ	ooooo x	oo	oo										

o - поверхностные царапины
x - раковины

• - неправильная форма
□ - другие

Δ - дефекты конечной обработки

**Таблица критических точек распределения χ^2 (хи-квадрат)
критерия Пирсона**

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6	0,103	0,051	0,02
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19	16,9	3,33	2,7	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,4	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25	7,26	6,26	5,23
16	32	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,9
22	40,3	36,8	33,9	12,3	11	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16	14,3
30	50,9	47	43,8	18,5	16,8	15

Приложение Е

Таблица для определения уровня несоответствий продукции

Значение C_p или C_{pk}	Уровень несоответствий продукции в процентах несоответствующих единиц продукции, %	Значение C_p или C_{pk}	Уровень несоответствий продукции в процентах несоответствующих единиц продукции, %
0,33	32,2	1,00	0,27
0,37	26,7	1,06	0,15
0,55	9,9	1,10	0,097
0,62	6,3	1,14	0,063
0,69	3,8	1,18	0,040
0,75	2,4	1,22	0,025
0,81	1,5	1,26	0,016
0,86	0,99	1,30	0,0096
0,91	0,64	1,33	0,0066
0,96	0,40		

Измерители качества и целевые показатели

Название измерителя	Обозначение	Смысл показателя	Формула расчета	Желаемые уровни и целевые значения
Уровень соответствия при приемке с первого предъявления	A_1	Показывает долю принятых изделий в результате выходного контроля при первом предъявлении от общего числа проконтролированных изделий	$A_1 = 1 - \frac{a_1}{N}$	Чем больше, тем лучше. Наилучшее значение – 1,0
Доля несоответствующей продукции, доля брака	R	Показывают долю отклоненных изделий при выходном контроле продукции от общего числа проконтролированных изделий	$R = \frac{r}{N}$	Чем меньше, тем лучше. Наилучшее значение – 0,0
Количество несоответствующей продукции на 1 млн. выпущенной (проконтролированной)	PPM	Смысл идентичен показателю R	$PPM = \frac{r}{N} \cdot 1.000.000$	Чем меньше, тем лучше. Наилучшее значение – 0,0
Количество несоответствий на 100 выпущенных изделий	D	Поскольку у одного изделия может быть много контролируемых характеристик, то число несоответствий и число несоответствующих изделий может не совпадать. И поэтому число несоответствий часто не	$D = \frac{\sum d_i}{N} \cdot 100$	Чем меньше, тем лучше. Наилучшее значение – 0,0

		совпадает с числом несоответствующих изделий		
Количество несоответствий на 1 млн. возможностей появления этих несоответствий	$DPMO$	Предположим у одного изделия 20 контролируемых характеристик ($k=20$), т.е. каждое изделие имеет 20 возможностей появления дефекта. Данный показатель показывает сколько дефектов будет обнаружено на 1 млн. возможностей появления дефекта (в данном случае 1 млн. возможностей соответствует 50.000 проверенных изделий)	$DPMO = \frac{\sum d_i}{N \cdot k} \cdot 1.000.000$	Чем меньше, тем лучше. Целевое значение – 3,0
Промежуточный уровень соответствия	A_{np}	Данный показатель характеризует вероятность прохождения j -ой стадии процесса без единого дефекта	$A_{np_j} = 1 - \frac{\sum d_i}{N}$	Чем больше, тем лучше. Наилучшее значение – 1,0
Сквозной уровень соответствия	$A_{скв}$	Данный показатель характеризует вероятность прохождения всех стадий процесса без единого дефекта	$A_{скв} = A_{np_1} \cdot A_{np_2} \cdot \dots \cdot A_{np_n}$	Чем больше, тем лучше. Наилучшее значение – 1,0
Нормализованный уровень соответствия	$A_{норм}$	Если мы хотим назначить целевое значение показателя A_{npj} , то мы можем рассчитать его на основе целевого показателя сквозного соответствия. Данный показатель – целевая вероятность пройти j -ую стадию процесса без единого дефекта	$A_{норм} = \sqrt[\text{step}]{A_{цел\ скв}}$	Определяется на основе целевого показателя сквозного уровня соответствия $A_{скв}$

единица «sigma»	sigma	Показывает сколько раз стандартное отклонение процесса вместится в пределы спецификаций. Определяется на основе показателя DPMO	По таблице DPMO	Чем больше, тем лучше. Целевое значение – 6,0
Контрольные карты (количество выходов за контрольные границы)	X, R, S, C, U, Np, p	Показывает стабилен ли процесс. Указывает на время появления особых причин, которые привели к потере стабильности процесса	Правило 3 сигм	Чем меньше, тем лучше (у идеального процесса в среднем 3 на 1 000 точек)
Индексы воспроизводимости	C_p, C_{pk}	Показывает на сколько реальный разброс процесса вписывается в поле допуска по данной контролируемой характеристике. Показывает насколько реальный разброс и смещение процесса способны выдавать годную продукцию	$C_p = \frac{U - L}{6\sigma}$ $C_{pk} = \min \left[\frac{U - \bar{X}}{3\sigma}; \frac{\bar{X} - L}{3\sigma} \right]$ $\sigma = \frac{\bar{R}}{d_2}$	Чем больше, тем лучше, но не менее 1,0. Целевое значение – не менее 1,33
Индексы пригодности	P_p, P_{pk}	Эти показатели имеют такой же смысл, что и C_p, C_{pk} . Они применяются для нестабильных процессов. Стандартное отклонение в данном случае считается по формуле отличной от формулы стандартного отклонения у C_p, C_{pk} .	$P_p = \frac{U - L}{6\sigma}$ $P_{pk} = \min \left[\frac{U - \bar{X}}{3\sigma}; \frac{\bar{X} - L}{3\sigma} \right]$ $\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$	Чем больше, тем лучше, но не менее 1,0. Целевое значение – не менее 1,33

Условные обозначения:

N – количество проконтролированных изделий за отчетный период

a_i – количество несоответствующих изделий при приемке с первого предъявления за отчетный период

r - количество несоответствующих изделий за отчетный период

d_i – количество несоответствий, обнаруженных у i -ой контролируемой характеристике за отчетный период

k – количество контролируемых характеристик за отчетный период

$step$ – количество технологических стадий

σ – стандартное отклонение

U, L – верхняя и нижняя границы допуска

\bar{x} - среднее значение контролируемой характеристики за определенный период времени

Таблица преобразований для 6 сигма

sigma	DPMO	sigma	DPMO	sigma	DPMO	sigma	DPMO
0,1	919 243	1,6	460 172	3,1	54 799	4,6	968
0,2	903 199	1,7	420 740	3,2	44 565	4,7	687
0,3	884 930	1,8	382 088	3,3	35 930	4,8	483
0,4	864 334	1,9	344 578	3,4	28 717	4,9	337
0,5	841 345	2,0	308 537	3,5	22 750	5,0	233
0,6	815 940	2,1	274 253	3,6	17 865	5,1	159
0,7	788 145	2,2	241 964	3,7	13 904	5,2	108
0,8	758 036	2,3	211 856	3,8	10 724	5,3	72
0,9	725 747	2,4	184 060	3,9	8 198	5,4	48
1,0	691 462	2,5	158 655	4,0	6 210	5,5	32
1,1	655 422	2,6	135 666	4,1	4 661	5,6	21
1,2	617 911	2,7	115 070	4,2	3 467	5,7	13
1,3	579 260	2,8	96 800	4,3	2 555	5,8	9
1,4	539 828	2,9	80 757	4,4	1 866	5,9	5
1,5	500 000	3,0	66 807	4,5	1 350	6,0	3