

**БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
ХАНТЫ-МАНСЙСКОГО АВТОНОМНОГО ОКРУГА – ЮГРЫ
«СУРГУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**



Система менеджмента качества

СМК СурГУ МИ-1.8.2-16

**Статистические методы
для анализа данных**

Редакция №1

стр. 1 из 27



УТВЕРЖДАЮ
Ректор С.М. Косенко

2016 г

Система менеджмента качества СурГУ. Методическая инструкция

СТАТИСТИЧЕСКИЕ МЕТОДЫ ДЛЯ АНАЛИЗА ДАННЫХ

МИ-1.8.2-16


СОГЛАСОВАНО:

Должность	Фамилия И.О.	Подпись	Дата
Первый проректор	Даниленко И.Н.		02.02.2016г.
Начальник Административно-правового управления	Бронников А.А.		02.02.2016г.

СОСТАВИЛ:

Начальник отдела менеджмента качества образования	Климович Л.А.		02.02.2016
---	---------------	--	------------

г. Сургут – 2016


	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 2 из 27

Содержание

1. Назначение и область применения	3
2. Нормативные ссылки	3
3. Термины и определения	3
4. Общие положения	5
5. Определение объема выборки.....	5
6. Контрольный листок и контрольная карта	6
7. Гистограмма.....	9
8. Стратификация (расслоение)	10
9. Диаграмма Исикавы	11
10. Диаграмма разброса	12
11. Диаграмма Парето.....	14
12. Анализ «Хи-квадрат».....	15
13. Среднее, медиана и мода	16
14. Перцентиль.....	18
15. Стандартное отклонение (разброс, изменчивость).....	20
Приложение 1. Таблица случайных чисел.....	21
Приложение 2. Пример составления контрольного листка	22
Приложение 3. Значения множителей для построения контрольных карт.....	23
Приложение 4. Значения «Хи-квадрат».....	24
Лист регистрации изменений.....	26
Лист ознакомления.....	27

Список используемых сокращений

- ДП – документированная процедура
МИ – методическая инструкция
ОРК – общее руководство по качеству

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 3 из 27

1. Назначение и область применения

1.1. Настоящая методическая инструкция (далее – МИ) определяет основные статистические методы, используемые для анализа данных в БУ ВО «Сургутский государственный университет» (далее – СурГУ, Университет).

1.2. Настоящая МИ обязательна к использованию всеми сотрудниками, осуществляющими анализ той или иной деятельности и (или) ее результатов.

2. Нормативные ссылки

2.1. МИ разработана на основе следующих нормативных документов:

- ГОСТ ISO 9000-2015 Системы менеджмента качества. Общие положения и словарь;
- ГОСТ ISO 9001-2015 Системы менеджмента качества. Требования;
- ГОСТ Р 52614.2-2006 Руководящие указания по применению ГОСТ Р ИСО 9001-2001 в сфере образования;
- ГОСТ Р 50.1.059-2006 Статистические методы. Руководство по выбору статистических методов для стандартов и технических условий;
- ГОСТ Р 50779.42-99 Статистические методы. Контрольные карты Шухарта;
- ОРК СурГУ Общее Руководство по качеству;
- ДП-1.7.1 «Управление документацией СМК»;
- ДП-1.7.2 «Управление записями».

3. Термины и определения

Анализ данных – деятельность, предпринимаемая для установления пригодности, адекватности, результативности рассматриваемого объекта для достижения установленных целей.

Блочная диаграмма – диаграмма, представляющая пять основных характеристик распределения.

Вероятность – степень возможности наступления каждого из множества потенциальных будущих событий, которая основывается на совокупности предположений.

Взвешенное среднее – сумма произведений значений элементов и их весов:

$$B_{\text{ср}} = \sum_{i=1}^n \omega_i X_i / n,$$

где $\omega_1, \omega_2, \dots, \omega_n$ – соответствующие веса, сумма которых равна единице, X_1, X_2, \dots, X_n – непосредственно сами значения данных, n – количество рассматриваемых объектов.

Выборка – множество случаев (испытуемых, объектов, событий, образцов), с помощью определённой процедуры выбранных из генеральной совокупности для участия в исследовании.


Выброс – точка, выпадающая из общего характера взаимосвязи в данных.

Генеральная совокупность – изучаемая совокупность некоторых единиц (людей, объектов и т.п.).

Гистограмма – изображение частот значений в виде совокупности столбцов, возвышающихся над числовой линией.

Диаграмма Парето – инструмент, позволяющий выявить и отобразить проблемы, установить основные факторы, с которых нужно начинать действовать, и распределить усилия с целью эффективного разрешения этих проблем.

Дисперсия – мера изменчивости для метрических данных, пропорциональная сумме квадратов отклонений измеренных значений от их арифметического среднего.

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 4 из 27

$$D = \frac{\sum_{i=1}^N (X_i - X_{cp})^2}{N},$$

Изменчивость – степень отличия значений данных друг от друга.

Квартили – это порядковые характеристики, отделяющие четверти ранжированных совокупностей (25-й и 75-й перцентили).

Контрольная карта – инструмент, позволяющий отслеживать изменение показателя качества во времени для определения стабильности процесса, а также корректировки процесса для предотвращения выхода показателя качества за допустимые пределы.

Контрольный листок – форма для регистрации и подсчета данных, собираемых в результате наблюдений или измерений контролируемых показателей в течении установленного периода времени.

Корреляция – степень зависимости между двумя переменными, изменение одной из них приводит к изменению другой.

Кумулятивная кривая (кумулянта) – ломаная, составленная по последовательно суммированным, т.е. накопленным частотам или относительным частотам.

Медиана – значение, расположенное посередине ряда данных таким образом, что делит данные на две половины.

Мода – наиболее распространенная категория; значение, чаще всего встречающееся в совокупности данных.

Нормальное распределение – непрерывное распределение, представленное кривой Гаусса (напоминающее колокол).

Нулевая гипотеза – это гипотеза о том, что две совокупности, которые сравниваются по одному или нескольким признакам, не отличаются.

Объём выборки — число случаев, включённых в выборочную совокупность.

Отклонение – расстояние между отдельным значением и средним.

Перцентиль – это сотая доля объема измеренной совокупности, выраженная в процентах.

Порядковые данные – данные, которые могут быть упорядочены.

Размах – результат вычитания наименьшего значения набора данных из наибольшего, характеризует размер или протяжённость набора данных.

Ранг – позиция значения в наборе данных после того, как набор упорядочен.

Репрезентативная выборка – выборка, в которой каждая характеристика наблюдается такой же процент раз, как и в соответствующей генеральной совокупности.


Систематизация — процедура объединения, сведения групп однородных по неким признакам единиц (параметрам, критериям) к определенному единству на основе существующих между ними связей.

Среднее – метод вычисления типичного значения для некоторой совокупности чисел путем сложения всех этих чисел и последующего деления полученной суммы на число элементов совокупности.

Стратификация (расслоение) – процесс сортировки (разделения) полученных данных на отдельные группы (страты) согласно некоторым критериям или факторам, результаты которого часто представлены в виде диаграмм или графиков.

Число степеней свободы – это количество независимых перемещений, при котором состояние системы меняется.

Функция кумулятивного распределения – функция, характеризующая распределение случайной величины или случайного вектора.

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 5 из 27

4. Общие положения

4.1. Анализ данных проводится при обработке результатов опроса, мониторинга и иных исследований деятельности Университета и (или) её результатов.

4.2. Перед исследованием, в зависимости от генеральной совокупности, определяется объем выборки в соответствии с разделом 5 настоящей МИ.

4.3. Для проведения полученных анализа данных могут использоваться следующие статистические методы или их комбинации:

- контрольный листок и контрольная карта;
- гистограмма;
- стратификация (расслоение);
- диаграмма Исикавы;
- диаграмма разброса;
- диаграмма Парето;
- анализ «Хи-квадрат»;
- среднее, меридиана и мода;
- перцентиль;
- стандартное отклонение (разброс, изменчивость).

4.4. Результатом анализа данных является отчет, содержащий обоснованные интерпретированные ответы (выводы), соответствующие поставленной цели исследования.

5. Определение объема выборки

5.1. Анализ данных может осуществляться как по результатам наблюдений за всеми контролируемыми объектами, так и выборки из общего числа объектов.

5.2. Объем (численность) выборки зависит от объема генеральной совокупности и цели анализа.

5.3. В зависимости от объема генеральной совокупности рекомендованы следующие объемы выборки из генеральной совокупности:

- при доверительной вероятности 95% (значение t-статистики Стьюдента – 2) (Таблица 1);
- при доверительной вероятности 90% (значение t-статистики Стьюдента – 2,75) (Таблица 2).

Таблица 1


Объем выборки в зависимости о генеральной совокупности

Объем генеральной совокупности	100	150	200	300	400	500	1000	2000	3000	4000	5000
Объем выборки	80	109	133	171	200	222	285	333	352	363	370

Таблица 2

Объем выборки в зависимости о генеральной совокупности

Объем генеральной совокупности	100	150	200	300	400	500	1000	2000	3000	4000	5000
Объем выборки	75	99	119	148	169	185	227	256	267	273	277

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 6 из 27

5.4. Доверительная вероятность 95% является предпочтительной. Если исследования проводятся при меньших объемах выборки, то необходимо определить меньшую доверительную вероятность результатов исследования (допустимую ошибку).

5.5. Рассматриваемая выборка должна быть репрезентативной. Пример нерепрезентативной выборки может быть в случае, когда в выборке доля студентов мужского пола больше, чем в генеральной совокупности, это означает, что выборка смещена в сторону студентов мужского пола.

5.6. Извлечение случайной выборки происходит двумя способами:

- автоматическим, с применением программы Microsoft Excel;
- вручную, с применением таблицы случайных чисел (Приложение 1).

Извлечение случайной выборки вручную осуществляется следующим способом:

- составление основы выборки (все элементы генеральной совокупности нумеруются от 1 до N);
- выбор точки начала считывания случайных чисел (Приложение 1);
- считывание цифры обычным способом, начиная с выбранной точки;
- объединение цифр в группы, размер которых равен количеству цифр в числе N ;
- получение выборки из n элементов.

Примером извлечения случайной выборки может быть 9 потребителей, из списка, в который входят 38 человек. Из таблицы случайных чисел (Приложение 1) выбираем точку начала считывания – число 69506. Число $N=38$ и состоит из двух цифр, значит необходимо объединить последовательность случайных чисел в группы, состоящие из двух чисел. Получаем (Таблица 3):

Таблица 3

Пример извлечения случайной выборки


№ п/п	Наименование этапа	Результат
1	Выбор точки начала считывания из таблицы случайных чисел	69506 19610 01479 92338 55140 81097 73071 61544 85356 51400
2	Объединение цифр в группы по две (т.к. $N=38$ и состоит из двух цифр)	69 50 61 96 10 01 47 99 23 38 55 14 08 10 97 73 07 16 15 44 85 35 65 14 00
3	Исключение чисел, которые больше 38 или меньше 1	10 01 23 38 14 08 10 07 16 15 35 14
4	Исключение чисел, которые уже встречались ранее	10 01 23 38 14 08 07 16 15 35
5	Выбор первых девяти чисел	10 1 23 38 14 8 7 16 15

Объекты, находящиеся под полученными числами и будут являться выборкой. После определения объема выборки и ее извлечения можно приступить к анализу.

6. Контрольный листок и контрольная карта

6.1. Методы контрольных листков и карт позволяют отслеживать состояние процесса во времени, а также воздействовать на процесс до того как он выйдет из-под контроля.

6.2. Контрольный листок позволяет ответить на вопрос: «Как часто случается определённое событие и на каком этапе?». Форма контрольного листка (пример см.

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 7 из 27

Приложение 2) разрабатывается в соответствии с конкретной ситуацией и может заполняться в электронном виде.

6.3. Построение контрольного листка включает в себя следующие шаги:

- определение событий (несоответствий, проблем), которые необходимо регистрировать в контрольном листке;
- определение периода сбора данных и подходящий интервал (может колебаться от часов до недель);
- определение категорий данных для регистрации в контрольном листке;
- инструктаж сотрудников, ответственных за сбор информации – как заполнять контрольный листок, какие события необходимо в нем регистрировать и в каком интервале времени;
- сбор данных по исследуемой проблеме и регистрация по каждому наблюдению (измерению) в соответствующей категории;
- суммирование данных по каждой категории и по всему интервалу наблюдений;
- обработка данных, анализ достижения поставленных целей и выводы.

6.4. Данные полученные в результате анализа с использованием контрольных листков, можно применить для дальнейшей работы по другим статистическим методам.

6.5. Контрольные карты используют для оценки: находится ли исследуемый процесс в контролируемом (управляемом) состоянии.

6.6. Контрольная карта, построенная на основании данных измерений показателей процесса в различные периоды времени и содержащая центральную линию и контрольные границы, график которой представляет собой ломаную линию и позволяет отразить динамику изменений показателя и за счет этого контролировать процесс. Существует два вида контрольных карт: \bar{X} -карта и \bar{R} -карта. На \bar{X} -карту наносят среднее каждой выборки, соответствующую центральную линию и контрольные границы, что позволяет наблюдать уровень процесса. На \bar{R} -карту наносят размах значений для каждой выборки, центральную линию и контрольные границы, что позволяет наблюдать изменчивость процесса.

6.7. Порядок построения контрольной карты:

- определение показателей процесса, подвергающиеся измерению, которые могут иметь количественные или качественные значения;
- определение точек контроля показателя;
- проведение измерений выбранного показателя процесса;
- отображение результатов измерений показателей на контрольной карте;
- соединение точек графика между собой;
- расчет центральной линии средних значений и вычисление контрольных границ, которые показывают выход из-под контроля процесса.
- отображение центральных линий средних значений и контрольных границ на контрольной карте;
- определение всех точек, выходящих за пределы контрольных границ.

6.8. Существует два способа определения контрольных границ и центральной линии в зависимости от наличия (отсутствия) стандартов (т.е. когда показатель численно ограничен (не ограничен)). Способы вычисления центральной линии и контрольных границ для \bar{X} -карты и \bar{R} -карты (Таблица 4).

Таблица 4

Способы вычисления центральной линии и контрольных границ

		Центральная линия	Контрольные границы
\bar{X} -карта	Стандарт задан (μ_0 и σ_0)	μ_0	От $\mu_0 - A\sigma_0$ до $\mu_0 + A\sigma_0$
	Стандарт не задан	$\bar{\bar{X}}$	От $\bar{\bar{X}} - A_2\bar{R}$ до $\bar{\bar{X}} + A_2\bar{R}$
\bar{R} -карта	Стандарт задан (σ_0)	$d_2\sigma_0$	От $M_1\sigma_0$ до $M_2\sigma_0$
	Стандарт не задан	\bar{R}	От $M_3\bar{R}$ до $M_4\bar{R}$

где μ_0 – заданное по стандарту значение показателя, σ_0 – заданное отклонение, $A, A_2, M_1, M_2, M_3, M_4$ – множители для контрольных границ (значения см. Приложение 3), \bar{R} – среднее значение размаха для всех выборок, d_2 – множитель для центральной линии (значения см. Приложение 3), $\bar{\bar{X}}$ – среднее значение всех выборочных средних.

Примером для построения \bar{X} - карты, когда стандарт не задан, может являться результат деятельности кафедр, рассматриваемых как одна выборка. Наибольшее значение выборки принимается равным удвоенному среднему значению, а наименьшее значение – нулю. В качестве рабочей величины для расчёта контрольных границ используется среднее каждой выборки (\bar{X} -карты). Так как стандарт значений не задан, центральная линия определяется как среднее значение $\bar{\bar{X}}$ всех выборочных средних \bar{X} . Контрольные границы располагаются симметрично относительно центральной линии.

6.9. Если значения показателей не выходят за пределы контрольных границ (рис.2.), то процесс считается контролируемым. Если значения показателя выходят за пределы контрольных границ (рис.3.), то процесс не находится под контролем и требует дальнейшего анализа для определения причин выхода значений показателей за пределы контрольных границ с применением других статистических методов. Необходимо обратить внимание и на те показатели, значения которых не вышли за пределы контрольных границ, но стремятся к ним.



Рис.2. Пример контрольной карты, когда процесс находится под контролем



Рис.3. Пример контрольной карты, когда процесс не находится под контролем

7. Гистограмма

7.1. Гистограмма позволяет оценить состояние качества и представляет собой столбиковую диаграмму частот, построенную по полученным за определенный период (час, неделю, месяц и т.д.) данным, которые разбиваются на несколько интервалов. Число данных, попавших в каждый из интервалов (частота), выражается высотой столбика.

7.2. Порядок составления гистограммы:

- сбор данных x_i контролируемого параметра за определенный период (час, неделю, месяц и т.д.). Число данных n должно быть не менее 30-50, оптимальное число порядка 100;
- определение наибольшего X_{max} и наименьшего X_{min} значения из всех полученных данных и вычисляется размах R :

$$R = X_{max} - X_{min},$$

который характеризует разброс контролируемой величины и определяет ширину гистограммы;

- определение количества интервалов по формуле Стерджесса:

$$N = 1 + 3.322Lgn,$$


Где n – объём, рассматриваемой выборки.

Рекомендуемое число интервалов гистограммы, которое получается при использовании формулы Стерджесса, представлено в Таблице 5.

Таблица 5

Рекомендуемое число интервалов на гистограмме

Количество данных в выборке	Число интервалов
23 — 45	6
46 — 90	7
91 — 180	8
181 — 361	9
362 — 723	10
724 — 1447	11
1448 — 2885	12

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 10 из 27

– определение ширины интервала:

$$h = R / N,$$

где R – размах, N – количество интервалов.

Все полученные данные распределяют по интервалам. Если какое-то значение попадает на границу, его следует относить к левому по отношению к ней интервалу. Подсчитывается число значений, попавших в каждый интервал N_i , где i -номер интервала.

– для каждого интервала подсчитывается относительная частота $F(x)$ попадания количество x_i , в рассматриваемый интервал.

По полученным данным строится гистограмма (рис.4), высота столбиков которой соответствует частоте или относительной частоте попадания данных в каждый из интервалов.

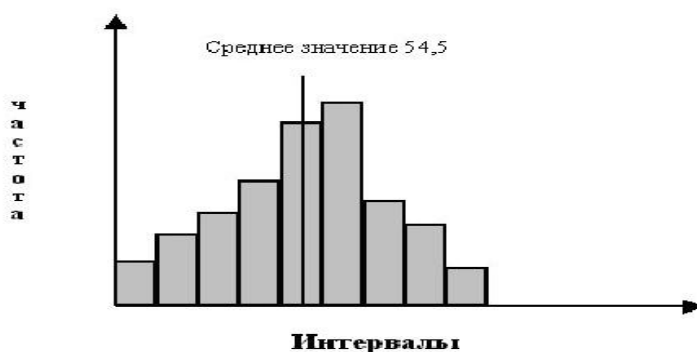


Рис.4. Гистограмма

7.3. Полученная в результате анализа гистограммы информация может быть легко использована для построения и исследования диаграммы Исикавы, что повысит обоснованность мер, намеченных для улучшения процесса.

8. Стратификация (расслоение)

8.1. С помощью стратификации происходит группировка общих данных проблем на подгруппы, что позволяет распознать наиболее проблемные участки.


8.2. Первый этап для проведения стратификации – сбор данных. Форма для сбора и учёта проблем разрабатывается в электронном виде. Возможно использование в качестве данных результаты анализа с использованием контрольных листов.

8.3. Обязательные поля формы сбора данных для стратификации:

- время проблемы;
- место проблемы;
- описание проблемы;
- причину проблемы;
- сотрудников, ответственных за проблему;
- действия по исправлению проблемы (коррекция и предупреждение);
- лиц, ответственных за проведение коррекции и предупреждения.

8.4. После накопления достаточного количества данных о проблемах происходит расслаивание по критериям (рис.5):

- процессам обслуживания;
- сотрудникам, участвовавшим в оказании услуги;
- оказанной услуге;
- потребителям.

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 11 из 27

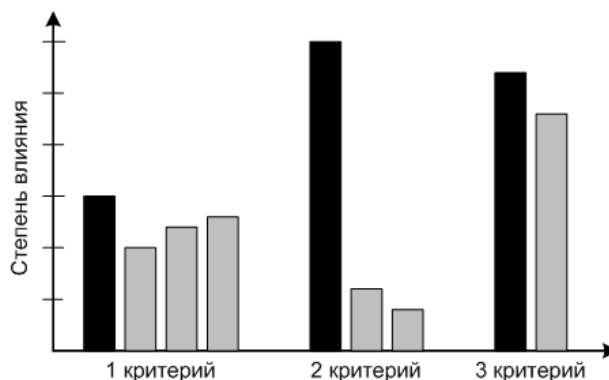


Рис.5. Пример стратификация по критериям

8.5. После проведения стратификация необходимо выделить из собранных результатов часто повторяющиеся группы данных (проблемные группы студентов, предметы, сотрудники, периоды, в которых чаще всего происходят проблемы и т.п.).

8.6. По результатам стратификации статистических данных могут быть сформулированы предложения по улучшению качества процесса.

9. Диаграмма Исикавы

9.1. Диаграмма Исикавы (причинно-следственная диаграмма) применяется для графического отображения взаимосвязи между решаемой проблемой и причинами, влияющими на ее возникновение. Данный инструмент используют совместно с методом «мозгового штурма», т.к. он позволяет быстро отсортировать по ключевым категориям причины проблем.

9.2. Диаграмма Исикавы строится следующим образом:

- определение потенциальной или существующей проблемы, требующей разрешения: формулировка проблемы размещается в прямоугольнике с правой стороны листа бумаги, от прямоугольника влево проводится горизонтальная линия (рис.6.);
- по краям листа с левой стороны обозначаются ключевые категории причин, влияющих на исследуемую проблему (количество и наименование категорий может изменяться в зависимости от рассматриваемой проблемы; как правило, используются пять или шесть следующих категорий: человек, методы работы, механизмы, материал, контроль, среда);
- от названий каждой из категорий причин к центральной линии проводятся наклонные линии, которые будут являться основными «ветвями» диаграммы Исикавы (рис. 7);
- причины проблемы распределяются по установленным категориям и указываются на диаграмме в виде «ветвей», примыкающих к основным «ветвям» (рис.8);
- каждая из причин детализируется на составляющие. Для этого по каждой из них задается вопрос – «Почему это произошло»? Результаты фиксируются в виде «ветвей» следующего, более низкого порядка. Процесс детализации причин продолжается до тех пор, пока не будет найдена «корневая» причина.

Проблема, которую
надо решить

Рис. 6. Первый этап построения диаграммы Исикавы

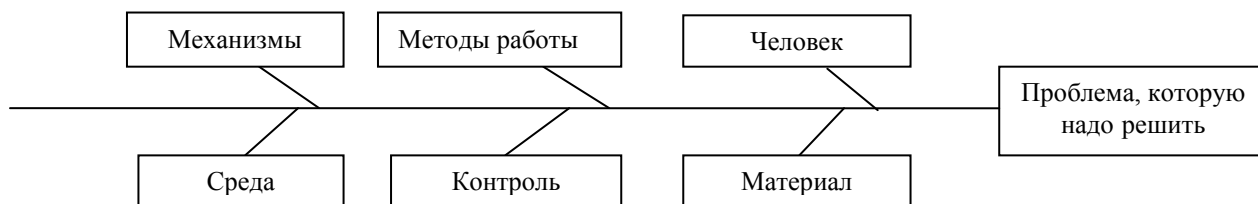


Рис.7. Второй этап построения диаграммы Исикавы

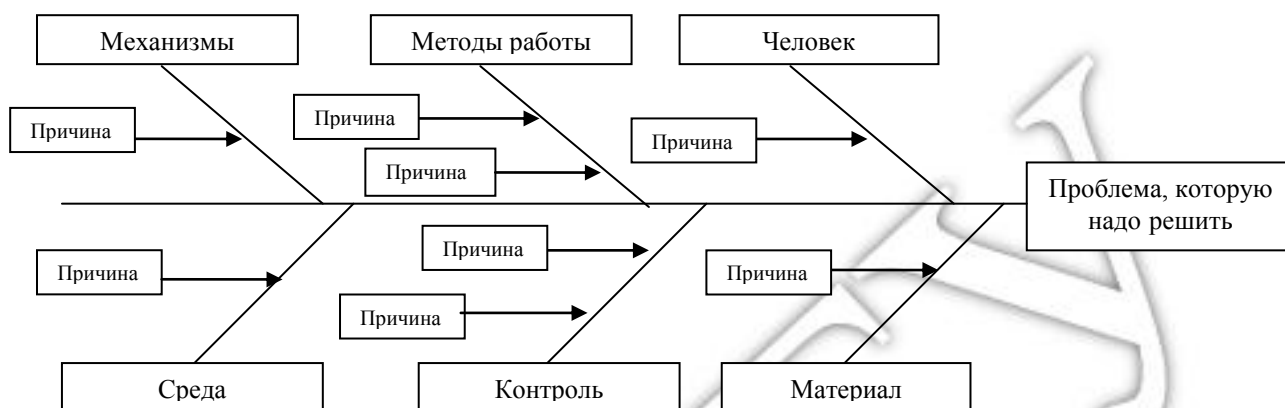


Рис. 8. Третий этап построения диаграммы Исикавы

9.3. При составлении причинно-следственной диаграммы последней стрелкой среди причин следует обозначить «и прочие неучтенные факторы».

9.4. Результаты детализации причин используются для выявления наиболее значимых и важных причин, влияющих на исследуемую проблему, например с помощью диаграммы Парето. По значимым причинам проводится дальнейшая работа, и определяются корректирующие или предупреждающие действия.

10. Диаграмма разброса

10.1. Диаграмма разброса предназначена для выявления зависимости между двумя типами данных и корреляции между каким-либо параметром качества и влияющим на него фактором.

10.2. Применяется диаграмма разброса в случае, когда необходимо отобразить что происходит с одной переменной при изменении другой, для определения причины возникновения неконтролируемых точек в ходе многовариантного статистического контроля процесса, подтверждения взаимосвязи, выявленной в результате применения причинно-следственной диаграммы (диаграммы Исикавы) и пр.

10.3. Построение диаграммы разброса включает в себя следующие шаги:

- сбор парных данных (не менее 20-25), которые по предположению являются взаимосвязанными;
- составление списка пар данных (Таблица 6);
- определение максимальных и минимальных значений по каждому из типов парных данных;
- выбор шкал для осей диаграммы разброса на основании разницы между максимальным и минимальным значением каждого из типов парных данных;

– построение горизонтальной (X) и вертикальной (Y) осей диаграммы. Шкала значений данных, обозначаемая на осях должна увеличиваться при подъеме по вертикальной оси и при движении вправо по горизонтальной. При исследовании корреляции между причиной и следствием (например, после применения диаграммы Исикавы) данные, характеризующие причину, откладываются по горизонтальной оси, а данные, характеризующие следствие – по вертикальной;

– нанесение на плоскость точек, имеющих координаты – парные данные. Если для разных измерений получаются одинаковые значения данных, то для отделения данных друг от друга используется другое обозначение (например, точки и треугольники) или данные обозначаются рядом друг с другом.

Таблица 6

Данные для построения диаграммы разброса

X	X_1	X_2	...	X_i	...	X_{n-1}	X_n
Y	Y_1	Y_2	...	Y_i	...	Y_{n-1}	Y_n

10.4. Форма и расположение точек на диаграмме разброса определяют различные варианты корреляции парных данных (рис.9).



Рис.9. Варианты корреляции парных данных

При прямой корреляции увеличение показателя одной переменной приводит прямо пропорциональному увеличению второй переменной. При отрицательной корреляции увеличение показателя одной из переменных приводит к уменьшению второй. Отсутствие корреляции указывает на то, что при изменении одной переменной нельзя спрогнозировать изменение другой, они независимы. При криволинейной корреляции одна переменная зависит от другой по не линейному уравнению.

10.5. После того как диаграмма разброса построена и в зависимости от того какая связь определена между двумя переменными проводится дальнейшая работа и определяются корректирующие или предупреждающие действия.

11. Диаграмма Парето

11.1. Диаграмма Парето позволяет разделить факторы, влияющие на возникшую проблему, на важные и несущественные для распределения усилий по ее решению.

11.2. В основе диаграммы Парето лежит принцип 20/80, согласно которому 20% всех причин являются существенными и приводят к 80% имеющихся проблем, поэтому целью построения диаграммы является выявление этих причин для концентрации усилий по их устранению.

11.3. Диаграмма Парето является разновидностью столбчатого графика с кумулятивной кривой, в которой факторы распределены в порядке уменьшения значимости (силы влияния на объект анализа).

11.4. Диаграмма Парето строится в следующем порядке:

- определение проблемы, которую необходимо решить и выбор временного интервала для изучения проблемы;
- выбор типа данных (фактор) для анализа, который наиболее полно сможет охарактеризовать проблему (например, несоответствия, объем потерь, затраты и пр.). Выбранный тип данных должен быть разбит на подтипы;
- определение единицы измерения, соответствующей типу данных (например, количество несоответствий, их частота, процент затрат и т.п.);
- сбор статистических данных и выполнение их систематизация (количество несоответствий для каждой из причин). Для сбора и регистрации данных можно применять другие инструменты качества, например контрольный листок. Систематизация статистических данных представляется в виде таблицы;
- расположение причин в порядке уменьшения частоты несоответствий;
- построение столбчатой диаграммы, на которой отмечаются подтипы данных (причины) и частота их проявления (рис.10.). В прямоугольной системе координат по горизонтали откладываются равные отрезки, соответствующие подтипам данных, а по вертикали отмечается величина этих данных в порядке по убыванию.
- вычисление и отображение на диаграмме линии суммарных значений (например, накопленных процентов);
- анализ полученных результатов для разработки необходимых действий по решению проблемы.

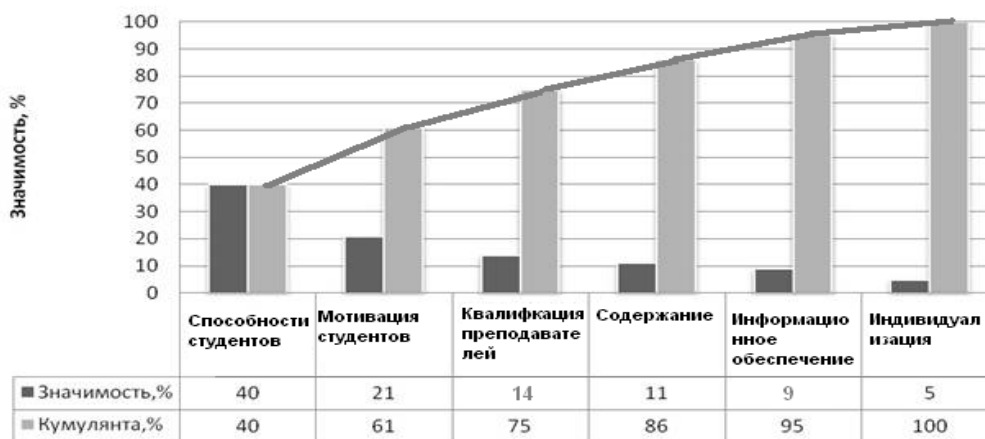



Рис.10. Пример построения диаграммы Парето

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 15 из 27

11.5. На диаграмме Парето самый высокий столбик будет всегда слева (показывая наиболее часто возникающие проблемы), а самый короткий справа. Линия, отражающая кумулятивный процент (кумулянту), всегда направлена вверх. Диаграмма Парето вносит объективность в вопросы о качестве и помогает сконцентрировать внимание на проблемах важных для Университета.

12. Анализ «Хи-квадрат»

12.1. Критерий «Хи-квадрат» используют для проверки гипотез о качественных данных, представленных не числами, а категориями. Критерий основан на частотах, которые представляют собой количество объектов выборки, попадающих в каждую из категорий.

12.2. «Хи-квадрат» измеряет разницу между наблюдаемыми (фактическими) частотами и ожидаемыми частотами по формуле:

$$\text{Хи - квадрат} = \sum \frac{O_i - E_i}{E_i}$$

где O_i - наблюдаемая частота, E_i – ожидаемая частота, Σ - сумма, вычисляется по всем категориям или комбинациям категорий.

12.3. Для обобщения качественных данных с помощью частот и процентов типичный набор данных представляется в виде списка результатов измерения для каждой из элементарных единиц выборки. Например, в качестве элементарных единиц могут выступать абитуриенты, пришедшие в Университет поступать на заочную форму обучения, тогда результатом измерения является предпочитаемое ими направление подготовки (наблюдаемое значение).

12.4. Перечень результатов измерения может быть очень длинным, поэтому рекомендуется работать с обобщающей таблицей частот или процентов (пример – Таблица 7).

12.5. «Хи-квадрат» используют в качестве меры того, насколько данные соответствуют нулевой гипотезе. Для того чтобы проверить эту гипотезу необходимо найти число степеней свободы по формуле:

$$Q = n - 1,$$

где Q – число степеней свободы, n – количество групп в списке.

В примере (Таблица 7) число групп равно 16, следовательно, число степеней свободы равно 15. По данным значения «Хи-квадрат» (Приложение 4) находим соответствующее значение. Уровень значимости, приведённый в приложении 4, определяет достоверность наблюдаемого результата, т.е. если уровень значимости равен 10 %, то наблюдаемый результат достоверен с вероятностью 90 %.

Таблица 7

Предпочитаемое направление подготовки

№ группы, n	Наименование направления подготовки	Результат подсчета (частота)	Процент от общего количества
1	Информатика и вычислительная техника	20	5,5
2	Управление в технических системах	9	2,5
3	Инфокоммуникационные технологии и системы связи	6	1,6
4	Электроэнергетика и электротехника	26	7
5	Строительство	30	8,2
6	Экономика	61	16,6

7	Менеджмент	26	7
8	Управление персоналом	36	10
9	Государственное и муниципальное управление	22	6
10	Юриспруденция	38	10,4
11	Физическая культура	17	4,6
12	Реклама и связи с общественностью	6	1,6
13	Экология и природопользование	15	4
14	Техносферная безопасность	17	4,6
15	Профессиональное обучение	38	10,3
16	Народная художественная культура	0	0
Итого		367	100

12.6. Если полученное значение по результатам вычисления «Хи-квадрат» больше, чем приведенное в Приложении 4, то это является свидетельством того, что наблюдаемые частоты значительно отличаются от тех, которые ожидаются. В этом случае следует отклонить нулевую гипотезу. Если полученное значение находится в пределах значений указанных в Приложении 4, то наблюдаемые значения не очень отличаются от значений, которые можно ожидать и нулевая гипотеза принимается.

12.7. Анализ «Хи-квадрат» можно также использовать при обработке данных полученных в результате тестирования.

13. Среднее, медиана и мода

13.1. Среднее, медиану и моду используют для выбора единственного числа (типического значения (центра)), которое лучше всего описывает все числа в наборе данных. В зависимости от типа данных определяют одну из описательных статистик (Таблица 8):

Таблица 8

Типы данных

№п/п	Описательная статистика	Количественные	Порядковые	Номинальные
1	Среднее	Да		
2	Медиана	Да	Да	
3	Мода	Да	Да	Да

13.2. Если распределение нормальное, то разница между показателями небольшая, поскольку каждая их статистик стремится к середине (рис. 11.).

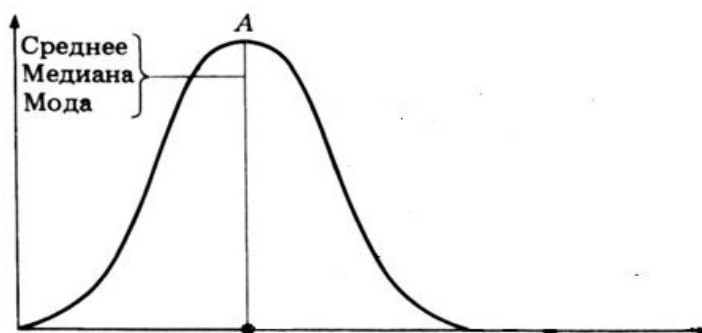



Рис. 11. Пример нормального распределения данных

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 17 из 27

13.3. Если распределение данных асимметричное значения среднего, медианы и моды могут заметно различаться (рис. 12.)

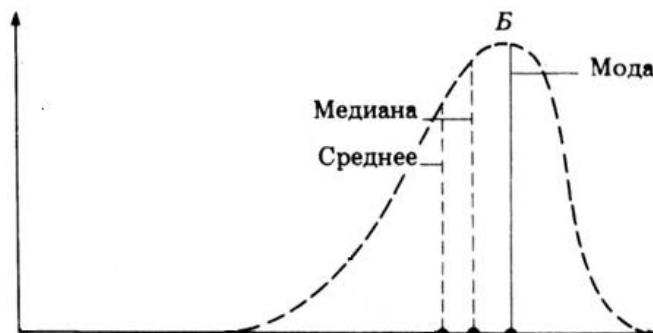


Рис. 12. Пример асимметричного распределения данных

13.4. Среднее следует использовать, когда набор данных распределен нормально, т.к. в этом случае среднее является самой эффективной характеристикой.

13.5. Медиану следует использовать при асимметричном распределении, т.к. на него не влияет небольшое число данных с высокими значениями.

13.6. Моду используют при наличии номинальных данных, т.к. в этом случае нельзя вычислять среднее и медиану. Также ее используют для порядковых данных, когда важно определить наиболее распространенную категорию.

13.7. Среднее можно вычислить только для имеющих содержательный смысл чисел (количественных данных) по формуле:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i,$$

где n – общее число элементов в списке данных, X_1, X_2, \dots, X_n – непосредственно сами значения данных, Σ – указывает на необходимость сложить все значения, которые записаны за ней, заменяя при этом индекс i значениями от 1 до n .


13.8. В случае, когда анализируемые данные не являются равноценными, используют взвешенное среднее (учет важности), оно дает возможность гибко определять систему важности отдельных элементов данных. Примером, когда применяют взвешенное среднее, может являться количество написанных монографий по трем различным институтам, где численность сотрудников может отличаться. Взвешенное среднее обобщает данные, используя веса, определенные в соответствии с количеством сотрудников в каждом институте. Для определения веса каждого института необходимо взять численность его сотрудников и разделить на общее количество сотрудников трёх институтов. Сумма полученных весов должна равняться единице.

Расчет взвешенного среднего осуществляется по формуле:

$$B_{зр} = \omega_1 X_1 + \omega_2 X_2 + \dots + \omega_n X_n / n,$$

где $B_{зр}$ – взвешенное среднее, $\omega_1, \omega_2, \dots, \omega_n$ – соответствующие веса, сумма которых равна единице, X_1, X_2, \dots, X_n – непосредственно сами значения данных, n – количество рассматриваемых объектов. Взвешенное среднее лучше всего интерпретировать как среднее, используемое в ситуациях, когда одни элементы более важны, чем другие.

13.9. Медиана (серединная точка) вычисляется как для порядковых данных, так и для чисел. Порядок вычисления медианы:

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 18 из 27

– расположение данных в порядке возрастания (убывания);

– определение среднего значения полученного ряда:

1) Если n - нечётное число, то медианой будет среднее значение данных, которое имеет номер $\lfloor \frac{n+1}{2} \rfloor$, если отсчитывать от любого из двух концов упорядоченного списка. Например, медиана списка 15, 27, 14, 18, 21 из $n=5$ значений равняется: *медиана (15,27,14,18,21) = медиана (14,15,18,21,27)*, 18 – это число третье по списку, что и соответствует формуле $\lfloor \frac{n+1}{2} \rfloor = 3$.

2) Если n – чётное число, то ряд имеет не одно, а два средних значения, которые расположены на расстоянии $\lfloor \frac{n+1}{2} \rfloor$ от каждого из двух концов упорядоченного списка данных. Например, медиана списка 15, 27, 14, 18 из $n = 4$ чисел вычисляется: *медиана (15,27,14,18) = медиана (14,15,18,27) = (15+18)/2 = 16,5*.

При определении среднего необходимо учитывать следующие условия:

– если набор данных количественный (состоит из чисел), то медианой является среднее этих двух значений, расположенных в середине ряда как на примере медианы списка 15, 27, 14, 18 равняется 16,5;

– если набор данных является порядковым (содержит упорядоченные категории) и если два расположенных в середине ряда значения представляют одну и ту же категорию, то эта категория является медианой, если различные категории, то обе эти категории будут медианами. Например, для списка рейтингов преподавателей обозначаемых A, AAA, B, AA, AAA , медиана вычисляется:

$$\text{медиана}(A, AAA, B, AA, AAA) = \text{медиана}(B, B, A, AA, AAA, AAA) = AA$$

Для порядковых данных нельзя вычислить среднее двух значений.

13.10. Мода вычисляется для номинальных данных, для порядковых данных и для чисел. Мода является наиболее распространенной категорией, которая чаще всего встречается в наборе данных.

13.11. Для количественных данных моду определяют как значение, соответствующее наивысшей точке на гистограмме, возможно, на середине самого высокого столбика.

13.12. Модой является категория с самым большим количеством или процентом в наборе данных. В случае если самыми большими являются две и больше категорий, то необходимо указывать все эти категории. Примером определения моды может быть голосование на выборах директора института. Все данные по результатам выборов представлены в Таблице 9.

Таблица 9


Результаты выборов

№ п/п	ФИО	Количество голосов	Процент
1	Иванов И.И.	20	33
2	Петров П.П.	16	26
3	Сидоров С.С.	25	41

Модой в этом наборе данных является Сидоров С.С., поскольку он набрал наибольшее количество голосов и наибольший процент.

14. Перцентиль

14.1. Перцентили рассматривают как показатели, разбивающие наборы количественных и порядковых данных на определенные части, выражающие ранги элементов в виде процентов от 0% до 100%, а не в виде чисел от 1 до n . Таким образом, наименьшему

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 19 из 27

значению соответствует нулевой перцентиль, наибольшему – 100-й перцентиль, медиане – 50-й перцентиль.

14.2. Перцентили используют для двух целей:

- чтобы показать значение элемента в данных при заданном перцентильном ранге;
- чтобы показать перцентильный ранг значения данного элемента в наборе данных.

14.3. Для установления соответствия между данными и процентами перцентилей используют функцию кумулятивного распределения, которая представляется в виде графика (рис. 13). График имеет вертикальный скачок величиной $1/n$ для каждого из n значений данных. По вертикальной оси откладываются проценты, а по горизонтальной сами перцентили (значения данных). Функция кумулятивного распределения отражает перцентили и позволяет их вычислить.

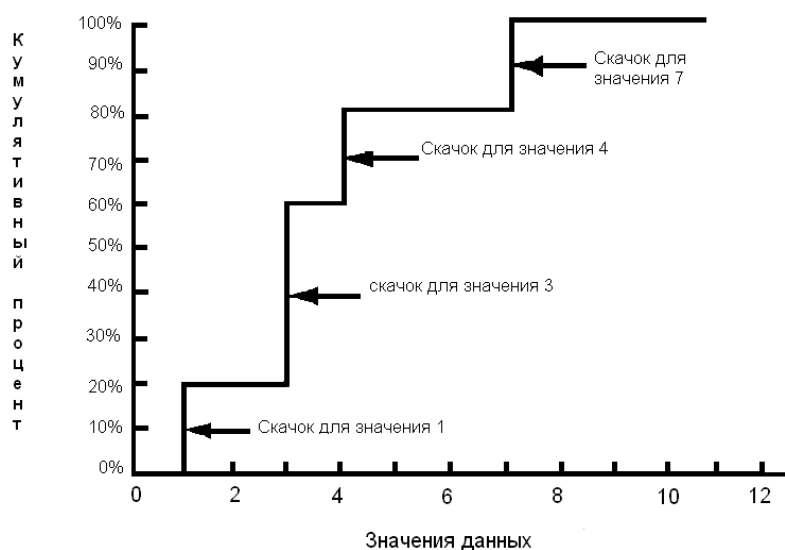


Рис. 13. Кумулятивное распределение и перцентили


Например (на рис. 13.) числу 4 соответствует 70-й перцентиль, так как перцентильный ранг этого значения расположен между 60 и 80%.

14.4. Одной из набора базовых характеристик являются квартили.

14.5. Порядок определения квартилей:

- вычисление ранга (места) медианы по формуле $\lfloor (n+1)/2 \rfloor$ без учёта дробной части. Например, при $n=13$ получаем $\lfloor (13+1)/2 \rfloor = 7$. При $n=24$ получаем $\lfloor (24+1)/2 \rfloor = 12,5$ без учёта дробной части равно 12;
- добавление к полученному значению единицы и деление на два. Полученное значение представляет собой ранг нижнего квартиля. Например, при $n=13$ ранг нижнего квартиля равен $\lfloor (7+1)/2 \rfloor = 4$. При $n=24$ ранг нижнего квартиля равен $\lfloor (12+1)/2 \rfloor = 6,5$, необходимо усреднить значения с рангами 6 и 7;
- вычитание полученного значения из $\lfloor (n+1)/2 \rfloor$. Результатом будет ранг верхнего квартиля. Например, при $n=13$ получается $\lfloor (13+1)/2 \rfloor - 4 = 10$. При $n=24$ получается $\lfloor (24+1)/2 \rfloor - 6,5 = 18,5$, необходимо усреднить значения с рангами 18 и 19.

14.6. Наименьшее значение (нулевой перцентиль), нижний квартиль, медиана (50-й перцентиль), верхний квартиль, наибольшее значение (100-й перцентиль) дают представление об особенностях еще не обработанного набора данных. Наибольшее и наименьшее значение характеризуют размах (диапазон) данных, медиана показывает

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 20 из 27

центр, два квартиля определяют границы, «расположенной в центре половины данных», а положение медианы относительно квартилей дает грубое представление о наличии или отсутствии асимметрии.

14.7. Для графического отображения пяти базовых показателей (наименьшее значение, нижний квартиль, медиана, верхний квартиль, наибольшее значение) используют блочную диаграмму (рис. 14), которая показывает точки (метки) выбросов (значения данных, расположенных далеко от центра распределения).

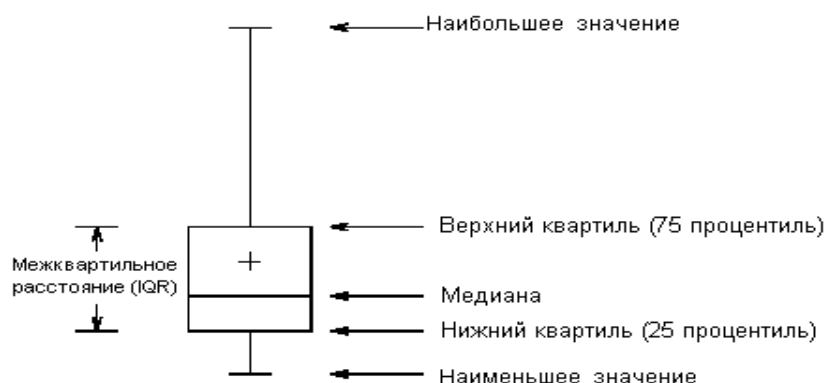


Рис. 14. Блочная диаграмма с пятью базовыми показателями

14.8. Большое значение в наборе данных рассматривается как выброс, если оно превышает: *верхний квартиль + 1,5 (верхний квартиль – нижний квартиль)*.

14.9. Малое значение в наборе данных рассматривается как выброс, если оно меньше чем: *нижний квартиль – 1,5 (верхний квартиль – нижний квартиль)*.

14.10. Метки выделяют те наблюдения, которые требуют особого внимания.

15. Стандартное отклонение (разброс, изменчивость)


15.1. Существуют три разных способа описания степени изменчивости набора данных:

- стандартное отклонение описывает, насколько сильно результат наблюдений обычно отличается от среднего значения;
- размах описывает, пределы изменения данных в наборе и представляет расстояние между минимальным и максимальным значениями;
- коэффициент вариации показывает, насколько сильно обычно отличается результат конкретного наблюдения от среднего значения, в процентном соотношении.

15.2. Стандартное отклонение является мерой случайности отклонений отдельных значений от их среднего. Порядок вычисления стандартного отклонения для выборки:

- определение отклонений, вычитая из каждого значения среднее (см. п.12.7.);
- возведение полученного значения в квадрат, после чего сложение и деление полученной суммы на $n-1$. Результатом будет дисперсия;
- извлечение из полученного значения квадратного корня.


15.3. Коэффициент вариации определяется как результат деления стандартного отклонения на среднее значение и выражается в процентах или долях среднего значения тем самым характеризует изменчивость. Коэффициент вариации показывает насколько отличается полученное значение от среднего.

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1
			стр. 21 из 27

Приложение 1

Таблица случайных чисел

№ п/п	1	2	3	4	5	6	7	8	9	10
1	51449	39284	85527	67158	91284	19954	91166	70918	85957	19492
2	16144	56830	67507	97275	25982	69294	32841	20861	83114	12531
3	48145	48280	99481	13050	81818	25282	66466	24461	97021	21072
4	83780	48351	85422	42978	26088	17869	94245	26622	48318	73850
5	95329	38482	93510	39170	63683	40587	80451	43058	81923	97072
6	11179	69004	34273	36062	26234	58601	47159	82248	95968	99722
7	94631	52413	31524	02316	27611	15888	13525	43809	40014	30667
8	64275	10294	35027	25604	65695	36014	17988	02734	31732	29911
9	72125	19232	10782	30615	42005	90419	32447	53688	36125	28456
10	16463	42028	27927	48403	88963	79615	41218	43290	53618	68082
11	10036	66273	69506	19610	01479	92338	55140	81097	73071	61544
12	85356	51400	88502	68267	73943	25828	38219	13268	09016	77465
13	84076	82087	55053	75370	71030	92275	55497	97123	40919	57479
14	76731	39755	78537	51937	11680	78820	50082	56068	36908	55399
15	19032	73472	79399	05549	14772	32746	38841	45524	13535	03113
16	72791	59040	61529	74437	74482	76619	05232	28616	98690	24011
17	11553	00135	28306	65571	34465	47423	39198	54456	95283	54637
18	71405	70352	46763	64002	62461	41982	15933	46942	36941	93412
19	17594	10116	55483	96219	85493	96955	89180	59690	82170	77643
20	09584	23476	09243	65568	89128	36747	63692	09986	47687	46448
21	81677	62634	52794	01466	85938	14565	79993	44956	82254	65223
22	45849	01177	13773	43523	69825	03222	58458	77463	58521	07273
23	97252	92257	90419	01241	52516	66293	14536	23870	78402	41759
24	26232	77422	76289	57587	42831	87047	20092	92676	12017	43554
25	87799	33602	01931	66913	63008	03745	93939	07178	70003	18158
26	46120	62298	69129	07862	76731	58527	39342	42749	57050	91725
27	53292	55652	11834	47581	25682	64085	26587	92289	41853	38354
28	81606	56009	06021	98392	40450	87721	50917	16978	39472	23505
29	67819	47314	96988	89931	49395	37071	72658	53947	11996	64631
30	50458	20350	87362	83996	86422	58694	71813	97695	28804	58523
31	59772	27000	97805	25042	09916	77569	71347	62667	09330	02152
32	94752	91056	08939	93410	59204	04644	44336	55570	21106	76588
33	01885	82054	45944	55398	55487	56455	56940	68787	36591	29914
34	85190	91941	86714	76593	77199	39724	99548	13827	84961	76740
35	97747	67607	14549	08215	95408	46381	12449	03672	40325	77312
36	43318	84469	26047	86003	34786	38931	34846	28711	42833	93019
37	47874	71365	76603	57440	49514	17335	71969	58055	99136	73589
38	24259	48079	71198	95859	94212	55402	93392	31965	94622	11673
39	31947	64805	34133	03245	24546	48934	41730	47831	26531	02203
40	37911	93224	87153	54541	57529	38299	65659	00202	07054	40168
41	82714	15799	93126	74180	94171	97117	31431	00323	62793	11995
42	82927	37844	74411	45887	36713	52339	68421	35968	67714	05883
43	65934	21782	35804	36676	35404	69987	52268	19894	81977	87764
44	56953	04356	68903	21369	35901	89797	83901	68681	02397	55359
45	16278	17165	67843	49349	90163	97337	35003	34915	91485	33814
46	96339	95028	48468	12279	81039	56531	10759	19579	00015	22829
47	84110	49661	13988	75909	35580	18426	29038	79111	56049	96451
48	49017	60748	03412	09880	94091	90052	43596	21424	16584	67970
49	43560	05552	54344	69418	01327	07771	25364	77373	34841	75927
50	25206	15177	63049	12464	16149	18759	96184	15968	89446	07168

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 22 из 27

Приложение 2

Пример составления контрольного листка

Контрольный листок

Цель: сбор данных по жалобам потребителей

Дата начала сбора данных: «16» сентября 2015 г.


Период сбора данных: сентябрь – декабрь 2015 г.

Ответственный за сбор данных: помощник ректора И.И. Иванов

№ п/п	Характер жалоб	Отметка (отмечайте черточкой каждую поступившую жалобу)	Частота
1	На преподавателя	Ш	3
2	На условия обучения	И	1
3	На работу кафедры	ШШ	4
4	На работу столовой	ШШШ	6
5	На отсутствие мест в общежитии	Ш	2

Значения множителей для построения контрольных карт


Размер выборки	Карты для средних (\bar{X} -карта)		Карты для диапазонов (\bar{R} -карта)				
	Множители для контрольных границ		Множитель для центральной линии	Множители для контрольных границ			
n	A	A_2	d_2	M_1	M_2	M_3	M_4
2	2,121	1,880	1,128	0	3,686	0	3,267
3	1,732	1,023	1,693	0	4,358	0	2,574
4	1,500	0,729	2,059	0	4,698	0	2,282
5	1,342	0,577	2,326	0	4,918	0	2,114
6	1,225	0,483	2,534	0	5,078	0	2,004
7	1,134	0,419	2,704	0,204	5,204	0,076	1,924
8	1,061	0,373	2,847	0,388	5,306	0,136	1,864
9	1,000	0,337	2,970	0,547	5,393	0,184	1,816
10	0,949	0,308	3,078	0,687	5,469	0,223	1,777
11	0,905	0,285	3,173	0,811	5,535	0,256	1,744
12	0,866	0,266	3,258	0,922	5,594	0,283	1,717
13	0,832	0,249	3,336	1,025	5,647	0,307	1,693
14	0,802	0,2235	3,407	1,118	5,696	0,328	1,672
15	0,775	0,223	3,472	1,203	5,741	0,347	1,653
16	0,750	0,212	3,532	1,282	5,782	0,363	1,637
17	0,728	0,203	3,588	1,356	5,820	0,378	1,622
18	0,707	0,194	3,640	1,424	5,856	0,391	1,608
19	0,688	0,187	3,689	1,487	5,891	0,403	1,597
20	0,671	0,180	3,735	1,549	5,921	0,415	1,585
21	0,655	0,173	3,778	1,605	5,951	0,425	1,575
22	0,640	0,167	3,819	1,659	5,979	0,434	1,566
23	0,626	0,162	3,858	1,710	6,006	0,443	1,557
24	0,612	0,157	3,895	1,759	6,031	0,451	1,548
25	0,600	0,153	3,931	1,806	6,056	0,459	1,541

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1
			стр. 24 из 27

Приложение 4

Значения «Хи-квадрат»

Число степеней свободы	Уровень значимости 10%	Уровень значимости 5%	Уровень значимости 1%	Уровень значимости 0,1%
1	2	3	4	5
1	2,706	3,841	6,635	10,828
2	4,605	5,991	9,210	13,816
3	6,251	7,815	11,345	16,266
4	7,779	9,488	13,277	18,467
5	9,236	11,071	15,086	20,515
6	10,645	12,592	16,812	22,458
7	12,017	14,067	18,475	24,322
8	13,362	15,507	20,090	26,124
9	14,684	16,919	21,666	27,877
10	15,987	18,307	23,209	29,588
11	17,275	19,675	24,725	31,264
12	18,549	21,026	26,217	32,909
13	19,812	22,362	27,688	34,528
14	21,064	23,685	29,141	36,123
15	22,307	24,996	30,578	37,697
16	23,542	26,296	32,000	39,252
17	24,769	27,587	33,409	40,790
18	25,989	28,869	34,805	42,312
19	27,204	30,144	36,191	43,820
20	28,412	31,410	37,566	45,305
21	29,615	32,671	38,932	46,797
22	30,813	33,924	40,289	48,268
23	32,007	35,172	41,638	49,728
24	33,196	36,415	42,980	51,179
25	34,382	37,652	44,314	52,620
26	35,563	38,885	45,642	54,052
27	36,741	40,113	46,963	55,476
28	37,916	41,337	48,288	56,892
29	39,087	42,577	49,588	58,301
30	40,256	43,733	50,892	59,703
31	41,422	44,985	52,191	61,098
32	42,585	46,194	53,486	62,487
33	43,745	47,400	54,776	63,870
34	44,903	48,602	56,061	65,247
35	46,059	49,802	57,342	66,619
36	47,212	50,998	58,619	67,985
37	48,363	52,192	59,893	69,346
38	49,513	53,384	61,162	70,703
39	50,660	54,572	62,428	72,055
40	51,805	55,758	63,691	73,402
41	52,949	56,942	64,950	74,745
42	54,090	58,124	66,206	76,084
43	55,230	59,304	67,459	77,419
44	56,369	60,481	68,710	78,749
45	57,505	61,656	69,957	80,077
46	58,641	62,830	71,201	81,400
47	59,774	64,001	72,443	82,720

	Сургутский государственный университет Система менеджмента качества	СМК СурГУ МИ-1.8.2-16	
		Статистические методы для анализа данных	Редакция №1 стр. 25 из 27

1	2	3	4	5
48	60,907	65,171	73,683	84,037
49	62,038	66,339	74,919	85,351
50	63,167	67,505	76,154	86,661
51	64,295	68,669	77,386	87,968
52	65,422	69,832	78,616	89,272
53	66,548	70,993	79,843	90,573
54	67,673	72,153	81,069	91,872
55	68,796	73,311	82,292	93,167
56	69,919	74,468	83,513	94,461
57	71,040	75,624	84,733	95,751
58	72,160	76,778	85,950	97,039
59	73,279	77,931	87,166	98,324
60	74,397	79,082	88,379	99,607
61	75,514	80,232	89,591	100,888
62	76,630	81,381	90,802	102,166
63	77,745	82,529	92,010	103,442
64	78,860	83,675	93,217	104,716
65	79,973	84,821	94,422	105,988
66	81,085	85,965	95,626	107,258
67	82,197	87,108	96,828	108,526
68	83,308	88,250	98,028	109,791
69	84,418	89,391	99,228	111,055
70	85,527	90,531	100,425	112,317
71	86,635	91,670	101,621	113,577
72	87,743	92,808	102,816	114,835
73	88,850	93,945	104,010	116,091
74	89,956	95,081	105,202	117,346
75	91,061	96,217	106,393	118,599
76	92,166	97,351	107,583	119,850
77	93,270	98,484	108,771	121,100
78	94,374	99,617	109,958	122,348
79	95,476	100,749	111,144	123,594
80	96,578	101,879	112,329	124,839
81	97,680	103,010	113,512	126,083
82	98,780	104,139	114,695	127,324
83	99,880	105,267	115,876	127,565
84	100,980	106,395	117,057	129,804
85	102,079	107,522	118,236	131,041
86	103,177	108,648	119,414	132,277
87	104,275	109,773	120,591	133,512
88	105,372	110,898	121,767	134,745
89	106,469	112,022	122,942	135,978
90	107,565	113,145	124,116	137,208
91	108,661	114,268	125,289	138,438
92	109,756	155,390	126,462	139,666
93	110,850	116,511	127,633	140,893
94	11,944	117,632	128,803	142,119
95	113,038	118,752	129,973	143,344
96	114,131	119,871	131,141	144,567
97	115,223	120,990	132,309	145,789
98	116,315	122,108	133,476	147,010
99	117,407	123,225	134,642	148,230
100	118,498	124,342	135,807	149,449

