

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
(ДГТУ)

Кафедра «Управление качеством»

МЕТОДЫ И ПРИЕМЫ ОПИСАТЕЛЬНОЙ СТАТИСТИКИ

Методические указания к практическим занятиям
по дисциплине «Статистическое управление»

Ростов-на-Дону
ДГТУ
2018

УДК 006.1

Составитель: О.Ю. Сорочкина

Методы и приемы описательной статистики: методические указания к практическому занятию по дисциплине «Статистическое управление» / Ростов н/Д, Издательский центр ДГТУ, 2018. - 12 с.

Приведены методы и инструменты описательной статистики, а также методика расчета основных показателей описательной статистики.

Методические указания предназначены для бакалавров направления подготовки 27.03.01 и магистров направления подготовки 27.04.01 Стандартизация и метрология, и бакалавров направления подготовки 27.03.02 Управление качеством.

УДК 006.1

Печатается по решению редакционно-издательского совета
Донского государственного технического университета

Научный редактор д-р техн. наук, профессор В.П. Димитров

Ответственный за выпуск зав. кафедрой «Управление качеством»
д-р техн. наук, профессор В.П. Димитров

В печать _____.____.2018 г.
Формат 60×84/16. Объем ____ усл. п. л.
Тираж ____ экз. Заказ №. ____.

Издательский центр ДГТУ
Адрес университета и полиграфического предприятия:
344000, г. Ростов-на-Дону, пл. Гагарина, 1

© Донской государственный технический университет, 2018

Введение

В практических наблюдениях мы обычно имеем совокупность наблюдений x_1, x_2, \dots, x_n , на основе которых требуется сделать те или иные выводы. Часто этих наблюдений много, поэтому возникает задача их компактного описания. В идеале таким описанием могло бы быть утверждение, что x_1, x_2, \dots, x_n являются выборкой, т.е. независимыми реализациями случайной величины ξ с известным законом распределения $F(x)$. Это позволило бы теоретически произвести расчеты всех необходимых исследователю характеристик наблюдаемого явления. [8]

Однако далеко не всегда мы можем утверждать, что x_1, x_2, \dots, x_n являются независимыми и одинаково распределенными случайными величинами. Во-первых, это необходимо проверить, а во-вторых, часто заведомо известно, что это не так. Поэтому для компактного описания совокупности наблюдений используют другие методы – методы описательной статистики. [3]

1. Методы описательной статистики

Методами описательной статистики называются методы описания выборок x_1, x_2, \dots, x_n с помощью различных показателей и графиков. Достоинство методов описательной статистики в том, что ее простые и довольно информативные статистические показатели избавляют от необходимости просмотра большого количества значений выборки. [3]

1 Показатели описательной статистики

Показатели, описывающие выборку можно разбить на несколько групп [4, 6]:

1. Показатели положения описывают положение данных (или середины совокупности) на числовой оси:

- Минимальный и максимальный элементы выборки
- Выборочные верхний и нижний квартили
- Среднее
- Выборочная медиана
- Выборочная мода

2. Показатели разброса описывают степень разброса данных относительно своего центра (насколько кучно основная масса данных группируется около середины совокупности)

- Дисперсия выборки
- Выборочное среднее квадратическое отклонение (стандартное отклонение)
- Размах
- Коэффициент эксцесса

3. Показатели асимметрии описывают симметричность распределения данных около своего центра

- Коэффициент асимметрии
- Положение выборочной медианы относительно выборочного среднего и относительно выборочных квартилей

- Гистограмма

4. Показатели, описывающие закон распределения, дают представление о законе распределения данных

- Гистограмма
- Выборочная функция распределения
- Таблица частот

Из перечисленных выше характеристик на практике по традиции чаще всего используют выборочные среднее, медиану и дисперсию (или стандартное отклонение).

Однако для получения более точных и достоверных выводов необходимо использовать и другие показатели.

Особое внимание следует обратить на наличие в выборке выбросов – грубых, сильно отличающихся от основной массы, наблюдений. Большинство традиционных статистических методов весьма чувствительны к отклонениям от условий применимости метода. Поэтому выбросы могут не только исказить значение выборочных показателей, но и привести к ошибочным выводам. Подозрение о присутствии таких наблюдений должно возникнуть, если выборочная медиана сильно отличается от выборочного среднего, хотя в целом совокупность симметрична, или, если положение медианы сильно несимметрично относительно минимального и максимального элементов выборки. Проще всего обнаружить выбросы с помощью перехода от выборки к вариационному ряду или гистограмме с большим числом интервалов группировки. [7]

2. Порядок выполнения работы

2.1. Исходные данные

Исходными данными является набор реализаций случайной величины (например, значения какой-либо величины, полученные при измерении). Размер выборки - n шт. Исходные данные оформить в виде таблицы (таблица 1).

Таблица 1 – Исходные данные

Номер реализации	Значение	Номер реализации	Значение	Номер реализации	Значение	Номер реализации	Значение
1...						...n	

2.2. Построение вариационного ряда

Для удобства работы с данными выборку преобразуют в вариационный ряд – ряд, в котором элементы выборки упорядочиваются по возрастанию.

Этапы выполнения:

1. Найти наименьший элемент ряда X_{\min}
2. Найти наибольший элемент ряда X_{\max}
3. Записать ряд, начиная с наименьшего элемента X_{\min} и заканчивая наибольшим X_{\max} (таблица 2)
4. Для упрощения процедуры обработки и с целью уменьшения ошибок при вычислениях необходимо вычесть из каждого элемента ряда постоянное число (например, округленное X_{\min}) и использовать в расчетах не сами размеры, а их отклонениями. Получившиеся отклонения записать в таблицу 2.

Таблица 2 – Вариационный ряд с отклонениями относительно $x_0 = \langle \text{значение} \rangle^1$

Номер элемента	Элемент	Отклонение	Номер элемента	Элемент	Отклонение
1...			n		

2.3. Группировка данных

Этапы выполнения:

1. Разбить весь диапазон $R = X_{\max} - X_{\min}$ на g интервалов. Число интервалов g устанавливают в зависимости от числа наблюдений n :

¹ - При оформлении работы выражения в угловых скобках заменить на конкретные значения

n	r
40-100	7-9
100-500	8-12
5000-10000	10-16

При небольших выборках [8]

$$r \approx \sqrt{n} . \quad (1)$$

2. Назначить длину интервалов. Длину интервалов Δx чаще всего выбирают одинаковой: $\Delta x = R/r$. Ее округляют до значения, удобного для графического отображения.

3. Назначить нижнюю границу x_n первого интервала (в отклонениях от x_0). Она должна быть меньше x_{\min} и удобной с позиции графического отображения. Результат занести в таблицу 3.

4. Назначить нижние x_n и верхние x_v границы всех оставшихся интервалов (в отклонениях от x_0). Результаты занести в таблицу 3.

5. Определить число размеров, попадающих в интервал m_i . Условие попадания размера x_j в интервал $x_n < x_j \leq x_v$. Результаты занести в таблицу 3. Полученные результаты проверить по условию:

$$\sum_{i=1}^r m_i = n . \quad (2)$$

2.4. Определение относительной частоты

Отношение частоты m_i к общему числу наблюдений n называется относительной частотой [7, 8]:

$$P_i^* = \frac{m_i}{n} \quad (3)$$

Относительная частота представляет собой эмпирическую оценку вероятности попадания результатов наблюдений X_j в i интервал.

Определить относительную частоту и результаты занести в таблицу 3.

Полученные результаты проверить по условию:

$$\sum_{i=1}^r P_i^* = 1 . \quad (4)$$

2.5. Определение эмпирической плотности вероятностей

Эмпирическая плотность вероятностей равна:

$$p_i^* = \frac{P_i^*}{\Delta x_i} \quad (5)$$

Определить эмпирическую плотность вероятности, результаты занести в таблицу 3.

Таблица 3 – Расчетные данные

Номера интервалов	Границы интервалов, <размерность>		Частота, m_i	Частость, P_i^*	Эмпирическая плотность вероятности p_i	Середина интервала x_i
	хн	хв				
1...						
			$\sum_{i=1}^r m_i =$	$\sum_{i=1}^r P_i^* =$		

Для дальнейших геометрических построений необходимы значения середины интервалов x_i . Определить их, результаты занести в таблицу 3.

2.6. Построение полигона

Этапы выполнения:

1. Определить масштабы по осям абсцисс и ординат, исходя их соотношения $P_i^*/R =$

5 ≈ 8.

На оси абсцисс отложить интервалы значений измеряемой величины.

2. В серединах интервалов отметить ординаты, пропорциональные частостям.

3. Полученные точки соединить прямыми линиями.

Пример полигона приведен на рисунке 1.

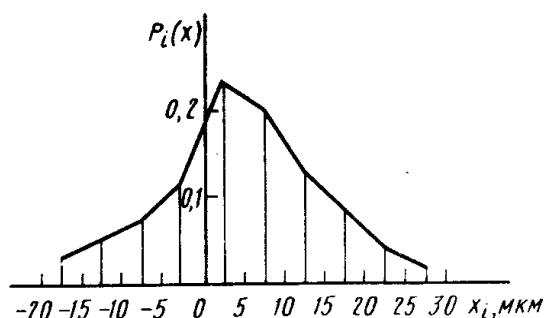


Рис. 1. Пример полигона

2.7. Построение гистограммы распределения

Этапы выполнения:

1. Повторить пункты 1-2 из 2.5.

2. Над каждым интервалом по оси абсцисс построить прямоугольник, высота которого пропорциональна эмпирической плотности вероятностей.

Пример гистограммы распределения приведен на рис. 2.

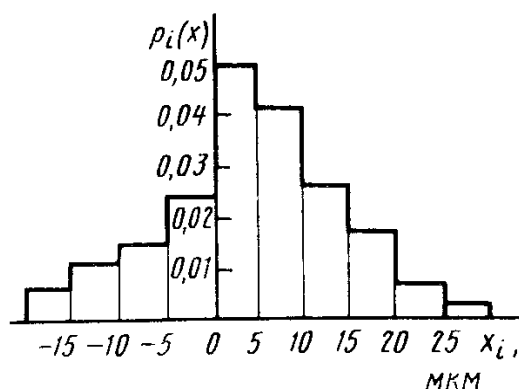


Рис. 2. Пример гистограммы распределения

2.8. Построение эмпирической функции распределения

В середине каждого интервала по оси абсцисс ордината возрастает скачком на значение, соответствующее P_i^* .

Этапы выполнения:

1. Повторить пункты 1-2 из 2.5.
2. В середине интервала 1 отметить скачок, равный P_1^* . Провести горизонтальную линию от получившейся точки до середины следующего интервала.
3. В середине интервала 2 отметить скачок от горизонтальной линии, полученной в п.2, равный P_2^* . Провести горизонтальную линию от получившейся точки до середины следующего интервала.
4. Повторить пункт 2 для остальных интервалов.

Значения F_i^* для каждого интервала называют накопленной относительной частотой, а сумму $\sum_{i=1}^i m_i$ - кумулятивной частотой.

Пример гистограммы эмпирической функции распределения приведен на рис. 3.

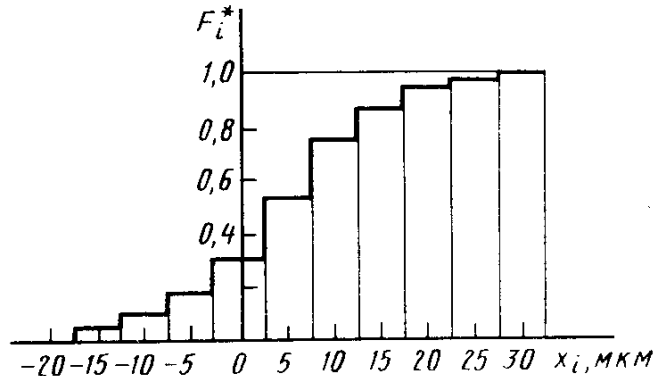


Рис. 3. Пример эмпирической функции распределения

2.9. Расчет параметров распределения

С помощью гистограммы распределения можно рассчитать параметры распределения [7, 8]:

1. Для среднего арифметического

$$\bar{X} = \sum_{i=1}^r X_i P_i^* = \sum_{i=1}^r X_i \frac{m_i}{n} = \frac{1}{n} \sum_{i=1}^r X_i m_i \quad (6)$$

2. Для выборочной дисперсии

$$S_X^2 = \sum_{i=1}^r (X_i - \bar{X})^2 P_i^* = \sum_{i=1}^r (X_i - \bar{X})^2 \frac{m_i}{n} = \frac{1}{n} \sum_{i=1}^r (X_i - \bar{X})^2 m_i \quad (7)$$

3. Для оценки центрального момента третьего порядка

$$m_3^* = \sum_{i=1}^r (X_i - \bar{X})^3 P_i^* = \sum_{i=1}^r (X_i - \bar{X})^3 \frac{m_i}{n} = \frac{1}{n} \sum_{i=1}^r (X_i - \bar{X})^3 m_i \quad (8)$$

4. Для оценки центрального момента четвертого порядка

$$m_4^* = \sum_{i=1}^r (X_i - \bar{X})^4 P_i^* = \sum_{i=1}^r (X_i - \bar{X})^4 \frac{m_i}{n} = \frac{1}{n} \sum_{i=1}^r (X_i - \bar{X})^4 m_i \quad (9)$$

Однако все расчеты можно значительно упростить, если все отклонения размеров y_i выражать относительными величинами в долях ширины интервала Δx (целыми числами), а за начало отсчета отклонений принять условный нуль x_0 , равный середине интервала, имеющего наибольшую частоту m_i :

$$y_i = \frac{x_i - x_0}{\Delta x} \quad (10)$$

Относительные начальные моменты в этом случае определяются:

$$a_j = \frac{1}{n} \sum_{i=1}^r y_i^j m_i \quad (11)$$

Возвращаясь к размерностям измеряемой величины, получим [1, 2]:

$$m_1 = \bar{x} = x_0 + a_1 \Delta x \quad (12)$$

$$m_2 = S^2 = \Delta x^2 (a_2 - a_1^2) \quad (13)$$

$$m_3 = \Delta x^3 (a_3 - 3a_2 a_1 + 2a_1^3) \quad (14)$$

$$m_4 = \Delta x^4 (a_4 - 4a_3 a_1 + 6a_2 a_1^2 - 3a_1^4) \quad (15)$$

$$\kappa_{ac} = \frac{m_3}{S^3} \quad (14)$$

$$\kappa_{экс} = \frac{m_4}{S^4} - 3 \quad (15)$$

Результаты расчета относительных начальных моментов удобнее всего свести в таблицу 4. [5]

Таблица 4 – Расчетные данные моментов

Номер интервалов	Середина интервала x_i , <размерность>	y_i	y_i^2	y_i^3	y_i^4	m_i	$m_i y_i$	$m_i y_i^2$	$m_i y_i^3$	$m_i y_i^4$
1										
...										
Суммы										
Начальные моменты										
Обозначения							a1	a2	a3	a4

Этапы выполнения:

1. Определить все произведения в таблице 4.
2. Определить все суммы в таблице 4.
3. Определить относительные начальные моменты в таблице 4.
4. Пересчитать полученные моменты для размерности измеряемой величины.
5. Рассчитать коэффициенты эксцесса и асимметрии.
6. Определить графически верхний и нижний квартили.
7. Определить графически выборочную медиану.
8. Определить выборочную моду (по вариационному ряду).
9. Определить размах.

2.10. Оформление результатов

Полученные результаты свести в таблицу 5 (нечисловые результаты заменить ссылками на рисунки и приложения). На рисунках в приложениях нанести (если возможно) параметры.

Таблица 5 – Показатели описательной статистики для выборки $n \leq >$

Показатель	Значение
1. Показатели положения	
– Минимальный элемент выборки MIN	
– Максимальный элемент выборки MAX	
– выборочный верхний квартиль $X_{0,75}$	
– выборочный нижний квартиль $X_{0,25}$	
– среднее \bar{x}	
– выборочная медиана M_d	
– выборочная мода M_o	
2. Показатели разброса	
– дисперсия выборки S^2	
– выборочное среднее квадратическое отклонение S	
– размах R	
– коэффициент эксцесса $k_{экс}$	
3. Показатели асимметрии	
– Коэффициент асимметрии $k_{ас}$	
– положение выборочной медианы относительно выборочного среднего и относительно выборочных квартилей	
– гистограмма	
4. Показатели, описывающие закон распределения	
– гистограмма	
– выборочная функция распределения	
– таблица частот	

3. Задачи для самостоятельного решения

Задача 3.1. При контроле массы готового изделия $m = 72,50$ г., были получены следующие данные в виде отклонений от номинального значения (таблица). Выполнить статистическую обработку данных.

№	Δm
1.	+ 3,34
2.	- 5,66
3.	+ 0,34
4.	+ 2,34
5.	+ 0,34
6.	+ 4,34
7.	+ 0,34
8.	+ 0,34
9.	+ 5,34
10.	- 1,66
11.	- 5,66
12.	- 1,66

№	Δm
13.	+13,34
14.	- 3,66
15.	- 0,66
16.	- 6,66
17.	+ 3,34
18.	+ 0,34
19.	- 6,66
20.	+ 2,34
21.	- 6,66
22.	- 4,66
23.	- 5,66
24.	+ 1,34

№	Δm
25.	+ 0,34
26.	+ 3,34
27.	- 1,66
28.	+ 0,34
29.	+2,34
30.	+ 2,34
31.	- 3,66
32.	+ 0,34
33.	- 5,66
34.	- 4,66
35.	+ 4,34
36.	+ 7,34

№	Δm
37.	+ 0,34
38.	+ 6,34
39.	+ 2,34
40.	- 3,66
41.	+ 9,34
42.	+ 7,34
43.	- 0,66
44.	- 1,66
45.	+ 4,34
46.	- 3,66
47.	- 5,66
48.	- 4,66

№	Δm
49.	+ 2,34
50.	- 6,66
51.	- 5,66
52.	+ 3,34
53.	- 4,66
54.	+ 2,34
55.	- 2,66
56.	+ 1,34
57.	+2,56
58.	-5,12
59.	-6,23
60.	+2,26

Задача 3.2. При контроле наружного диаметра радиального шарикового сферического

двухрядного подшипника (тип 1000 , № 1008) , были получены следующие данные в виде отклонений (мкм) от номинального значения $\varnothing 22$ (таблица). Выполнить статистическую обработку данных.

№	ΔD
1.	-4
2.	-2
3.	-5
4.	-6
5.	-3
6.	-9
7.	-2
8.	-1
9.	-4
10.	-8
11.	0
12.	0

№	ΔD
13.	-9
14.	-1
15.	-4
16.	-6
17.	-2
18.	-9
19.	0
20.	-9
21.	-3
22.	-4
23.	-7
24.	+1

№	ΔD
25.	0
26.	-2
27.	-8
28.	-7
29.	-6
30.	-7
31.	0
32.	+2
33.	-7
34.	-2
35.	-9
36.	0

№	ΔD
37.	-4
38.	-4
39.	-3
40.	-5
41.	-8
42.	-6
43.	-2
44.	-2
45.	-5
46.	-9
47.	-4
48.	-2

№	ΔD
49.	0
50.	0
51.	-5
52.	-3
53.	2
54.	-9
55.	-9
56.	-4
57.	-4
58.	-2
59.	-5
60.	0

Задача 3.3. Ниже показан вес 100 резиновых изделий, которые были последовательно произведены на одном и том же станке. Приведенные числа — это отклонения веса от номинала (в граммах). Выполнить статистическую обработку данных.

3	-3	-2	2	-4	-6	-2	-5	-2	5	5	3	2	6	2	4	7	11	6	3
3	-2	3	-1	-3	4	-1	-2	2	3	-4	-11	-5	-7	-5	-6	15	-5	-1	-1
1	3	7	-6	2	-1	-5	1	2	1	2	2	-1	-8	-7	5	3	6	13	1
-2	-2	4	9	1	6	7	1	3	2	5	2	9	5	-3	-3	-5	-1	-2	-1
3	-1	-1	1	-2	-3	2	-7	-9	-1	-3	-2	-6	-2	-6	-8	-2	4	-1	-1

Задача 3.4. Раскаленный металл транспортируется из доменной печи в сталеплавильную печь по рельсам. Время транспортировки (в минутах) для 140 последовательных разливок приведено ниже — по 20 значений в строке. Выполнить статистическую обработку данных.

40	45	125	100	40	40	100	65	55	40
120	45	65	105	35	70	55	25	50	55
40	70	40	40	110	55	50	30	50	105
55	45	100	60	45	145	45	50	65	180
45	35	45	55	50	70	45	75	60	45
50	25	25	100	50	60	45	35	40	30
65	65	45	40	50	25	120	115	50	85
125	65	40	45	95	105	45	110	40	50
50	40	40	45	55	50	45	125	55	100
45	45	55	50	25	65	60	60	55	70
60	45	35	35	55	55	55	50	120	35
60	40	60	40	50	60	65	95	65	60

180	50	30	30	30	65	130	80	20	45
40	35	40	40	30	55	50	25	75	55

Задача 3.5. При контроле внутреннего диаметра подшипника роликового цилиндрического для букс железнодорожного подвижного состава (тип 232000, № 232532), были получены следующие данные в виде отклонений (мкм) от номинального значения $\varnothing 290$ (таблица).

Выполнить статистическую обработку данных.

№	Δd
1.	-25
2.	-32
3.	-12
4.	-30
5.	-25
6.	-29
7.	-16
8.	-34
9.	-25
10.	-27
11.	-10
12.	-29

№	Δd
13.	0
14.	-30
15.	0
16.	0
17.	-25
18.	-29
19.	-16
20.	-18
21.	-28
22.	-34
23.	-29
24.	-38

№	Δd
25.	0
26.	-12
27.	-32
28.	-28
29.	-36
30.	-28
31.	-31
32.	-30
33.	-26
34.	-37
35.	-23
36.	-19

№	Δd
37.	-34
38.	-26
39.	-16
40.	-18
41.	-19
42.	-24
43.	-30
44.	-16
45.	0
46.	-19
47.	-25
48.	-29

№	Δd
49.	-26
50.	-16
51.	-27
52.	-30
53.	-29
54.	-34
55.	-20
56.	-25
57.	-16
58.	-19
59.	-25
60.	-27

Задача 3.6. При контроле выборки, были получены следующие результаты (таблица). Выполнить статистическую обработку данных.

	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
значение	70	72	71	70	69	68	73	70	69	69
	<i>11</i>	<i>12</i>	<i>13</i>	<i>14</i>	<i>15</i>	<i>16</i>	<i>17</i>	<i>18</i>	<i>19</i>	<i>20</i>
значение	68	62	70	73	72	70	70	71	75	70
	<i>21</i>	<i>22</i>	<i>23</i>	<i>24</i>	<i>25</i>	<i>26</i>	<i>27</i>	<i>28</i>	<i>29</i>	<i>30</i>
значение	70	69	62	70	70	73	74	70	70	75
	<i>31</i>	<i>32</i>	<i>33</i>	<i>34</i>	<i>35</i>	<i>36</i>	<i>37</i>	<i>38</i>	<i>39</i>	<i>40</i>
значение	72	70	69	69	64	70	71	71	73	70
	<i>41</i>	<i>42</i>	<i>43</i>	<i>44</i>	<i>45</i>	<i>46</i>	<i>47</i>	<i>48</i>	<i>49</i>	<i>50</i>
значение	69	70	65	70	73	75	70	71	71	72

Список использованной литературы

1. ГОСТ Р 50779.21-2004 Статистические методы. Правила определения и методы расчета статистических характеристик по выборочным данным. Часть 1. Нормальное распределение. - М.: ИПК Издательство стандартов, 2004.
2. ГОСТ Р 51814.3-2001 Системы качества в автомобилестроении. Методы статистического управления процессами. - М.: ИПК Издательство стандартов, 2001.
3. ГОСТ Р ИСО/ТО 10017-2005 Статистические методы. Руководство по применению в соответствии с ГОСТ Р ИСО 9001 - М.: Стандартиформ, 2005.
4. Вентцель Е.С. Теория вероятностей и ее инженерные приложения - М.: Высшая школа, 2000. — 480 с.
5. Вуколов Э. А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXSEL : учеб. Пособие - М.: Форум, 2008. – 464с.
6. Ефимов, В.В. Статистические методы в управлении качеством: Учебное пособие / В.В. Ефимов. - Ульяновск: УлГТУ, 2003. - 134 с.
7. Практикум по метрологии: пособие по решению задач/ И.Г. Кошлякова, В.А. Ваганов, А.Ф. Хлебунов. – Ростов н/Д: Издательский центр ДГТУ, 2009. – 214с.
8. Шишкин И.Ф. Теоретическая метрология, Часть 1, Общая теория измерений, / Учебник для вузов, 4-е издание, перераб. и доп.- Спб.: Питер, 2010. - 192 с. - ISBN:978-5-49807-203-6