

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Локтионова Оксана Геннадьевна

Должность: проректор по учебной работе

Дата подписания: 27.01.2023 22:54:50

Уникальный программный ключ:

0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

МИНОБРАЗОВАНИЯ И НАУКИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Юго-Западный государственный университет» (ЮЗГУ)

Кафедра космического приборостроения и систем связи

УТВЕРЖДАЮ

Проректор по учебной работе

О.Г. Локтионова

« 27 » января 2017 г.



**ПОСТРОЕНИЕ ЭМПИРИЧЕСКИХ ФУНКЦИИ И
ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ**

Методические указания к практическим занятиям
по дисциплине «Теоретические основы конструирования, техноло-
гии и надежности электронных средств» для бакалавров направле-
ния подготовки 11.03.03 «Конструирование и технология элек-
тронных средств»

Курск – 2017

УДК 681.3

Составитель Е. О. Брежнева

Рецензент

Доктор технических наук, профессор кафедры вычислительной техники *И.Е. Чернецкая*

Построение эмпирических функции и плотности распределения: методические указания к практическим занятиям по дисциплине «Теоретические основы конструирования, технологии и надежности электронных средств» / Юго-Зап. гос. ун-т; сост. Е.О. Брежнева. Курск, 2017. - 17 с.

Излагаются краткие сведения о формах представления эмпирических распределений, методике построения статистических рядов и их графическом представлении. Приведены варианты заданий и контрольные вопросы.

Методические указания соответствуют требованиям программы, утвержденной учебно-методическим объединением по специальностям автоматике и электроники (УМО АЭ).

Предназначены для магистров направления подготовки 11.03.03 «Конструирование и технология электронных средств».

Текст печатается в авторской редакции

Подписано в печать . Формат 60×84 1/16.

Усл. печ. л. 0,99. Уч.- изд. л. 0,89. Тираж 30 экз. Заказ . Бесплатно.

Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94

ОГЛАВЛЕНИЕ

1. Цель работы.....	4
2. Основные теоретические сведения.....	5
3. Задачи для решения.....	13
4. Контрольные вопросы.....	17

1. ЦЕЛЬ РАБОТЫ

- 1) Приобретение навыков построения статистических рядов.
- 2) Знакомство с графическим представлением статистических рядов.

2. ОСНОВНЫЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

2.1 Формы представления эмпирических распределений

Статистическое исследование состоит из трех основных этапов. Первым этапом является наблюдение, при котором производится научно обоснованный сбор данных, характеризующих изучаемое явление или объект. Вторым этапом является статистическая сводка и группировка. На этом этапе данные систематизируются и определенным образом оформляются – чаще всего в виде статистических таблиц. Третьим этапом является анализ статистического материала.

Применение того или иного метода статистического анализа определяется математической моделью, описывающей свойства генеральной совокупности. Для корректного проведения эксперимента выбор и обоснование математической модели должны быть произведены до его начала. На практике при проведении обследований ограниченный объем предварительной информации не позволяет сделать обоснованное предположение о математической модели генеральной совокупности. В таких случаях ее выбор осуществляется на основе построения эмпирического распределения и анализа его характеристик. Под **эмпирическим распределением** принято понимать распределение элементов выборки по значениям изучаемого признака. Построение эмпирических распределений является необходимым этапом применения статистических методов. Основной задачей при построении эмпирического распределения является формулирование на основе его анализа предположения о форме распределения изучаемого признака в генеральной совокупности.

2.2 Построение статистических рядов

Выборка, полученная при проведении экспериментального исследования, представляет собой неупорядоченный набор чисел, записанных в той последовательности, в которой производились измерения. Обычно выборка оформляется в виде таблицы, в первой

строке (или столбце) которой стоит номер опыта i , а во второй (втором) - зафиксированное значение случайной величины признака (таблица 1). В таком виде выборка представляет собой первичную форму записи статистического материала, который может быть обработан различными способами.

Таблица 1

№	1	2	3	4	5	6	7	8	9	10
x_i										

Проанализировать экспериментальные данные и тем более сделать какие-либо выводы на основе простой статистической совокупности весьма затруднительно. Исходя из этого, полученный статистический материал должен быть обработан для проведения дальнейшего исследования. Простейшим способом обработки выборки является ранжирование. Ранжированием называют расстановку вариантов в порядке возрастания или убывания их значений.

Но и в таком виде полученные экспериментальные данные плохо обозримы и мало пригодны для непосредственного анализа. Именно поэтому для придания статистическому материалу большей компактности и наглядности он должен быть подвергнут дальнейшей обработке – строится статистический ряд. Построение статистического ряда начинается с группировки.

Группировкой называется процесс упорядочения и систематизации данных, полученных в ходе проведения эксперимента, направленный на извлечение содержащейся в них информации. В процессе группировки осуществляется распределение вариантов выборки по группам или интервалам группировки, каждый из которых содержит некоторый диапазон значений изучаемого признака. Процесс группировки начинается с разбиения всего диапазона варьирования признака на интервалы группировки.

Ориентировочное значение оптимального числа интервалов k может быть определено, исходя из объема выборки n , с помощью формулы Стэрджесса:

$$k = 1 + 3,322 \lg n.$$

Получаемое по формуле значение k почти всегда оказывается дробной величиной, которую необходимо округлить до целого числа, поскольку количество интервалов не может быть дробным.

Практика показывает, что, как правило, лучше округлять в меньшую сторону.

После того, как число интервалов группировки определено, вычисляется ширина интервалов:

$$h = \frac{x_{\max} - x_{\min}}{k}.$$

Здесь h - ширина интервалов, а x_{\max} и x_{\min} - соответственно максимальное и минимальное значение признака в выборке. Величины x_{\max} и x_{\min} определяются непосредственно по таблице исходных данных.

Точность вычисленного значения h может либо совпадать с точностью проведения эксперимента, либо превышать ее. Особо отметим, что округление необходимо производить не в общепринятом математическом смысле, а в сторону увеличения, т.е. с избытком, чтобы не уменьшить общий диапазон варьирования признака - сумма ширины всех интервалов не должна быть меньше разности между максимальным и минимальным значениями признака.

После определения ширины интервалов группировки следует определить их границы. Нижнюю границу первого интервала целесообразно принять равной минимальному значению признака в выборке x_{\min} :

$$x_{H1} = x_{\min}.$$

Для получения верхней границы первого интервала (x_{B1}) следует к значению нижней границы первого интервала прибавить значение ширины интервала:

$$x_{B1} = x_{H1} + h.$$

Заметим, что верхняя граница каждого интервала (здесь - первого) будет являться одновременно и нижней границей следующего (в данном случае второго) интервала: $x_{H2} = x_{B1}$.

Подобным образом определяются значения нижних и верхних границ всех оставшихся интервалов:

$$x_{Bi} = x_{Hi+1} = x_{Hi} + h.$$

Перед группировкой вариант введем понятие **срединного значения интервала** x_i , равного значению признака, равноудаленного от концов этого интервала. Учитывая, что оно отстоит от

нижней границы на величину, равную половине ширины интервала, для его определения удобно воспользоваться соотношением:

$$x_i = x_{Hi} + h/2,$$

где x_{Hi} - нижняя граница i -го интервала, а h - его ширина. Срединные значения интервалов будут использоваться в дальнейшем при обработке сгруппированных данных.

После определения границ всех интервалов следует распределить выборочные варианты по этим интервалам. Если варианта может быть отнесена к любому из двух соседних интервалов, для исключения неоднозначности при группировке, условимся относить варианты к верхнему интервалу.

Перейдем к рассмотрению статистической таблицы (таблица 2.1), которая состоит из семи столбцов.

Таблица 2.1- Статистическая таблица

№ интервала	Границы интервала	Срединное значение интервала	Частота	Накопленная частота	Частость	Накопленная частость
i	$x_{Hi} - x_{Bi}$	x_i	n_i	N_i	f_i	F_i
			Σ		Σ	

В первых трех столбцах статистической таблицы содержатся соответственно номера интервалов группировки i , их границы $x_{Hi} - x_{Bi}$ и срединные значения интервалов x_i .

В четвертом столбце располагаются частоты интервалов. **Частотой** интервала называется число, показывающее сколько вариантов, т.е. результатов измерений попало в данный интервал. Для обозначения этой величины принято использовать символ n_i . Сумма всех частот всех интервалов всегда равна объему выборки n , что можно использовать для проверки правильности проведенной группировки.

Пятый столбец таблицы 2.1 предназначен для занесения в него **накопленной частоты** интервала - числа, полученного суммированием частоты текущего интервала с частотами всех предыдущих интервалов. Накопленную частоту принято обозначать ла-

тинской буквой N_i . Накопленная частота показывает, сколько вариантов имеют значения не больше, чем верхняя граница интервала.

В шестой столбец таблицы помещается частость. **Частостью** называется частота, представленная в относительном выражении, т.е. отношение частоты к объему выборки. Сумма всех частостей всегда равна 1. Для обозначения частости используется символ f_i :

$$f_i = n_i/n.$$

И частота, и частость характеризуют повторяемость результатов в выборке. Следует отметить важное достоинство частости. Ее использование позволяет сопоставлять выборки различного объема. Частота для таких целей не применима.

В седьмом столбце таблицы расположена накопленная частость. **Накопленной частостью** является отношение накопленной частоты к объему выборки. Накопленная частость обозначается буквой F_i :

$$F_i = \frac{N_i}{n}.$$

Накопленная частость показывает, какая доля вариантов выборки имеет значения, не превосходящие значения верхней границы интервала.

Последняя строка статистической таблицы используется для контроля над проведением группировки.

Статистическим рядом называется двойной числовой ряд, устанавливающий связь между численным значением исследуемого признака и его повторяемостью в выборке. Существенным достоинством статистических рядов является то, что они, в отличие от статистических совокупностей, дают наглядное представление о характерных особенностях варьирования признаков.

2.3 Графическое представление статистических рядов

В целях упрощения анализа статистических рядов и придания им большей наглядности используют графические представления. Основными видами графического представления статистических рядов являются гистограмма, полигон частостей и полигон накопленных частостей. Для визуального представления можно

использовать как частоты, так и частоты. Ограничимся рассмотрением частоты, поскольку этот параметр более информативен.

Наиболее часто для анализа статистического ряда используется **гистограмма**, представляющая собой совокупность примыкающих друг к другу прямоугольников, основание каждого из которых равно ширине интервала группировки, а площадь - частоты этого интервала.

Гистограмма строится в декартовой (прямоугольной) системе координат следующим образом. По оси абсцисс откладываются отрезки, отображающие интервалы группировки, а затем на каждом из них строится прямоугольник, площадь которого равна частоте данного интервала. Для удовлетворения этому требованию высота прямоугольника выбирается равной частному от деления частоты интервала на его ширину $H_i = f_i/h_i$. В случае, если все интервалы группировки имеют одинаковую ширину, высоты прямоугольников пропорциональны соответствующим частотам. Полная площадь гистограммы равна единице, что следует из способа ее построения. Действительно, площадь каждого из прямоугольников равна частоте, а сумма всех частотей - единица.

В качестве примера на рисунке 1 приведена гистограмма распределения.

С увеличением числа экспериментальных данных можно использовать большее количество интервалов, имеющих меньшие ширины. Гистограмма при этом будет все более и более приближаться к некоторой кривой, ограничивающей площадь, равную единице. Эта кривая представляет собой не что иное как график плотности распределения (или, по-другому, плотности вероятности) исследуемой случайной величины. Таким образом, гистограмма является экспериментальным аналогом плотности распределения.

Другим распространенным способом графического представления статистических рядов является полигон частотей. **Полигон частотей** отображает зависимость частоты от срединных значений интервалов. Полигон частотей строится в декартовой системе координат путем соединения прямыми линиями точек, абсциссы которых равны срединным значениям интервалов, а ор-

динаты - частостям этих интервалов. Эти данные располагаются в третьем и шестом столбцах таблицы 2.1.

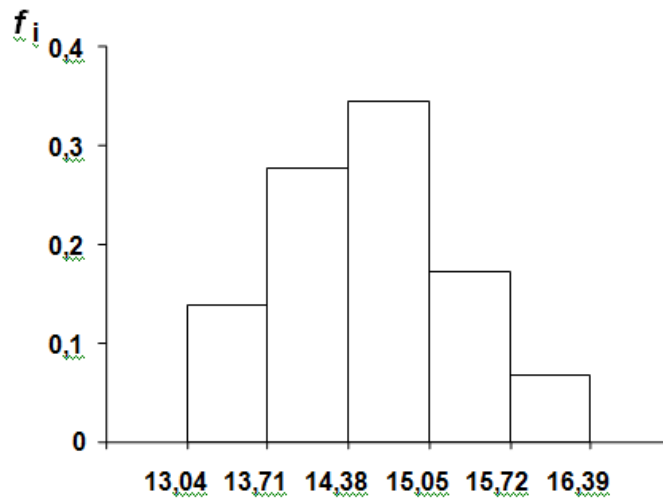


Рисунок 1 - Гистограмма

Полигон частостей может быть получен из гистограммы путем соединения середин верхних сторон прямоугольников гистограммы отрезками прямых. Полигон частостей изображен на рисунке 2.1.

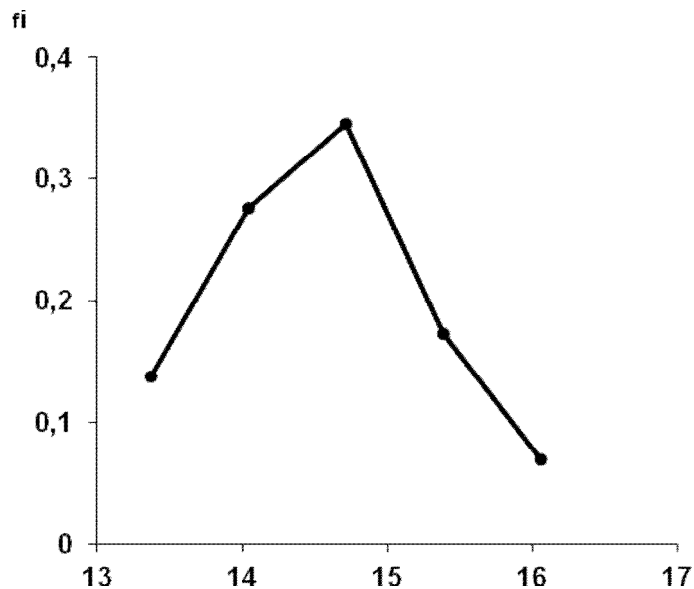


Рисунок 2 - Полигон частостей

Полигон частостей может оказаться более удобным и наглядным способом графического представления, чем гистограмма, в том случае, когда признак является непрерывным и его распределение описывается плавной зависимостью.

Полигон накопленных частот представляет собой зависимость накопленных частот от значений верхних границ интервалов. Полигон накопленных частот строится в декартовой системе координат посредством соединения прямыми линиями точек, абсциссы которых равны значениям верхних границ интервалов, а ординаты - накопленным частотам этих интервалов. Эти данные располагаются во втором и седьмом столбцах таблицы 2.1. Полигон накопленных частот приведен на рисунке 3.

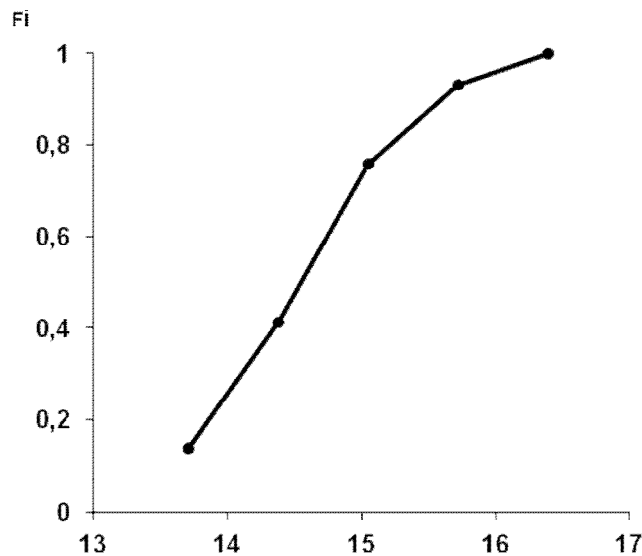


Рисунок 3. Полигон накопленных частот

С увеличением числа опытных данных в выборке и соответственно увеличением числа используемых интервалов полигон накопленных частот будет приближаться к кривой, являющейся графиком функции распределения исследуемой случайной величины. Таким образом, он является экспериментальным аналогом функции распределения.

3. ЗАДАЧИ ДЛЯ РЕШЕНИЯ

Для экспериментальных данных, в соответствии с номером варианта, построить статистический ряд, гистограмму, полигон частостей и полигон накопленных частостей. Сделать выводы.

Таблица 3.1- Вариант №1

№	1	2	3	4	5	6	7	8	9	10
x_i	0,53	1,83	-2,25	0,86	0,31	-1,30	-0,43	0,34	3,57	2,76
№	11	12	13	14	15	16	17	18	19	20
x_i	-1,34	3,03	0,72	-0,06	0,71	-0,20	-0,12	1,48	1,40	1,41
№	21	22	23	24	25	26	27	28	29	30
x_i	0,67	-1,20	0,71	1,63	0,48	1,03	0,72	-0,30	0,29	-0,78

Таблица 3.2- Вариант №2

№	1	2	3	4	5	6	7	8	9	10
x_i	101	99	99	100	98	102	101	100	102	99
№	11	12	13	14	15	16	17	18	19	20
x_i	100	100	101	101	100	100	100	101	102	102
№	21	22	23	24	25	26	27	28	29	30
x_i	100	101	99	99	100	102	100	101	100	102

Таблица 3.3- Вариант №3

№	1	2	3	4	5	6	7	8	9	10
x_i	29,2	10,0	25,4	26,3	27,3	11,6	17,9	15,1	26,0	18,6
№	11	12	13	14	15	16	17	18	19	20
x_i	28,2	13,6	15,2	12,9	12,7	27,3	21,5	20,9	12,8	27,0

№	21	22	23	24	25	26	27	28	29	30
x_i	22,4	17,0	20,2	18,0	11,5	14,7	12,4	13,6	14,7	18,3

Таблица 3.4- Вариант №4

№	1	2	3	4	5	6	7	8	9	10
x_i	12	47	48	30	30	24	47	25	15	42
№	11	12	13	14	15	16	17	18	19	20
x_i	26	20	27	14	16	48	49	34	13	20
№	21	22	23	24	25	26	27	28	29	30
x_i	25	43	11	12	17	36	40	36	29	32

Таблица 3.5- Вариант №5

№	1	2	3	4	5	6	7	8	9	10
x_i	14	19	13	18	13	15	17	19	11	21
№	11	12	13	14	15	16	17	18	19	20
x_i	19	16	15	15	14	16	16	19	19	18
№	21	22	23	24	25	26	27	28	29	30
x_i	15	19	16	14	21	20	17	17	17	13

Таблица 3.6- Вариант №6

№	1	2	3	4	5	6	7	8	9	10
x_i	0,52	0,15	0,29	0,18	0,57	0,69	0,07	0,10	0,02	0,89
№	11	12	13	14	15	16	17	18	19	20
x_i	0,68	0,92	0,28	0,65	0,04	0,92	1,24	0,23	0,10	0,46
№	21	22	23	24	25	26	27	28	29	30
x_i	0,68	0,43	0,62	0,26	0,03	0,18	0,91	0,37	1,16	0,27

Таблица 3.7- Вариант №7

№	1	2	3	4	5	6	7	8	9	10
x_i	12	11	15	11	12	11	26	15	12	16
№	11	12	13	14	15	16	17	18	19	20
x_i	17	12	13	13	12	24	15	14	17	12
№	21	22	23	24	25	26	27	28	29	30
x_i	11	16	12	19	11	19	13	15	11	14

Таблица 3.8- Вариант №8

№	1	2	3	4	5	6	7	8	9	10
x_i	14	11	15	14	11	24	11	18	12	13
№	11	12	13	14	15	16	17	18	19	20
x_i	12	15	13	11	28	21	11	12	17	14
№	21	22	23	24	25	26	27	28	29	30
x_i	20	16	16	15	19	15	20	11	21	11

Таблица 3.9- Вариант №9

№	1	2	3	4	5	6	7	8	9	10
x_i	5,68	4,14	3,92	4,90	4,74	6,19	5,60	5,54	3,55	4,03
№	11	12	13	14	15	16	17	18	19	20
x_i	5,20	4,65	6,29	6,34	4,41	5,87	6,39	5,32	6,62	6,06
№	21	22	23	24	25	26	27	28	29	30
x_i	5,21	5,87	5,19	4,58	5,35	5,03	4,63	6,77	5,22	7,73

Таблица 3.10- Вариант №10

№	1	2	3	4	5	6	7	8	9	10
x_i	15	17	20	21	18	13	18	11	15	18

№	11	12	13	14	15	16	17	18	19	20
x_i	21	19	16	19	15	21	21	20	15	16
№	21	22	23	24	25	26	27	28	29	30
x_i	13	19	20	21	17	17	12	20	15	13

4. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Дать определение случайной величины.
2. Генеральная и выборочная совокупности.
3. Формы представления эмпирических распределений.
4. Последовательность построения статистических рядов.
5. Определение числа и ширины интервалов.
6. Частота и накопленная частота интервалов: определения, расчет.
7. Частость и накопленная частость: определения, расчет.
8. Графическое представление статистических рядов. Гистограмма.
9. Графическое представление статистических рядов. Полигон частостей.
10. Графическое представление статистических рядов. Полигон накопленных частостей.