§8. Основные понятия математической статистики

п.1. Предмет математической статистики.

Математическая статистика изучает закономерности, которым подчинены массовые случайные явления, с помощью методов ТВ.

Пусть по результатам наблюдений изучается некоторая СВ.

Основные задачи мат. статистики:

— упорядочить исходные данные (представить их в виде, удобном для анализа);

— оценить требуемые характеристики наблюдаемой СВ (функцию распределения, мат. ожидание, дисперсию и т.д.);

— проверить статистические гипотезы, т.е. решить вопрос согласования результатов оценивания с данными.

п.2. Выборочный метод.

Совокупность всех подлежащих исследованию объектов называется генеральной совокупностью.

Выборочной совокупностью или случайной выборкой называют совокупность случайно отобранных объектов.

Объемом совокупности (выборочной или генеральной) называют число объектов этой совокупности.

Пусть в выборке событие наблюдалось раз, событие — раз и т.д., событие — раз.

Тогда объем выборки равен

Наблюдаемые значения называют **вариантами**.

Числа называют **частотами**.

Числа

называют **относительными частотами**.

Вариационным рядом называют таблицу вида:

X			$\ldots$	
n			$\ldots$	

Статистическим распределением выборки (**статистическим рядом**) называют таблицу вида:

X			$\ldots$	
w			$\ldots$	

Пример. В результате тестирования группа студентов получила следующие оценки

Построить вариационный и статистический ряд.

Решение. Вариационный ряд.

X						
n						

Статистический ряд:

X						
n						

Если число вариант велико или наблюдаемая СВ является непрерывной, то поступают следующим образом.

Вместо значений в первую строку вариационного (статистического) ряда записывают

(формула Стерджеса)

полуинтервалов длиной

Полученный ряд называют интервальным.

Пример.

189	207	213	208	186	210	198	219	231	227
202	211	220	236	227	220	210	183	213	190
197	227	187	226	213	191	209	196	202	235
211	214	220	195	182	228	202	207	192	226
193	203	232	202	215	195	220	233	214	185
234	215	196	220	203	238	225	221	193	215
204	184	217	193	216	205	197	203	229	204
225	216	233	223	208	204	207	182	216	191
210	190	207	205	232	222	198	217	211	201
185	217	225	201	208	211	189	205	207	199

Записать интервальный вариационный ряд.

Решение.

кол-во интервалов

длина интервала

Интервальный вариационный ряд:

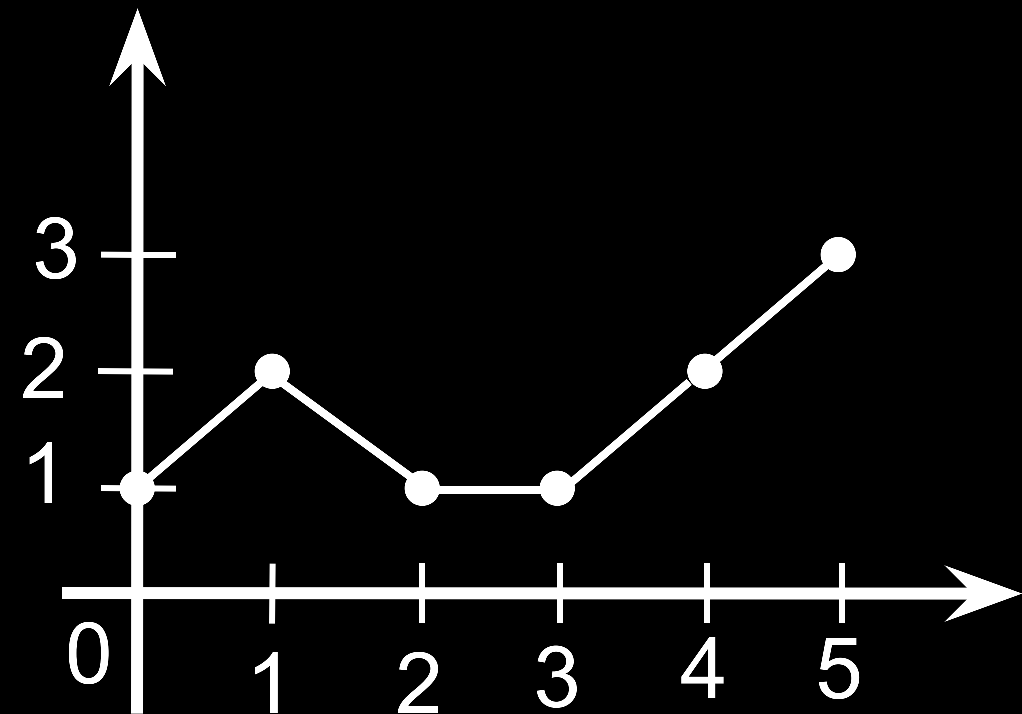
п.3. Полигон и гистограмма.

Полигоном частот называется ломаная, соединяющая точки с координатами

Пример. Вариационный ряд

X	0	1	2	3	4	5
n	1	2	1	1	2	3

Полигон частот:



Для изучения непрерывного признака строится гистограмма.

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников с основанием h и высотой

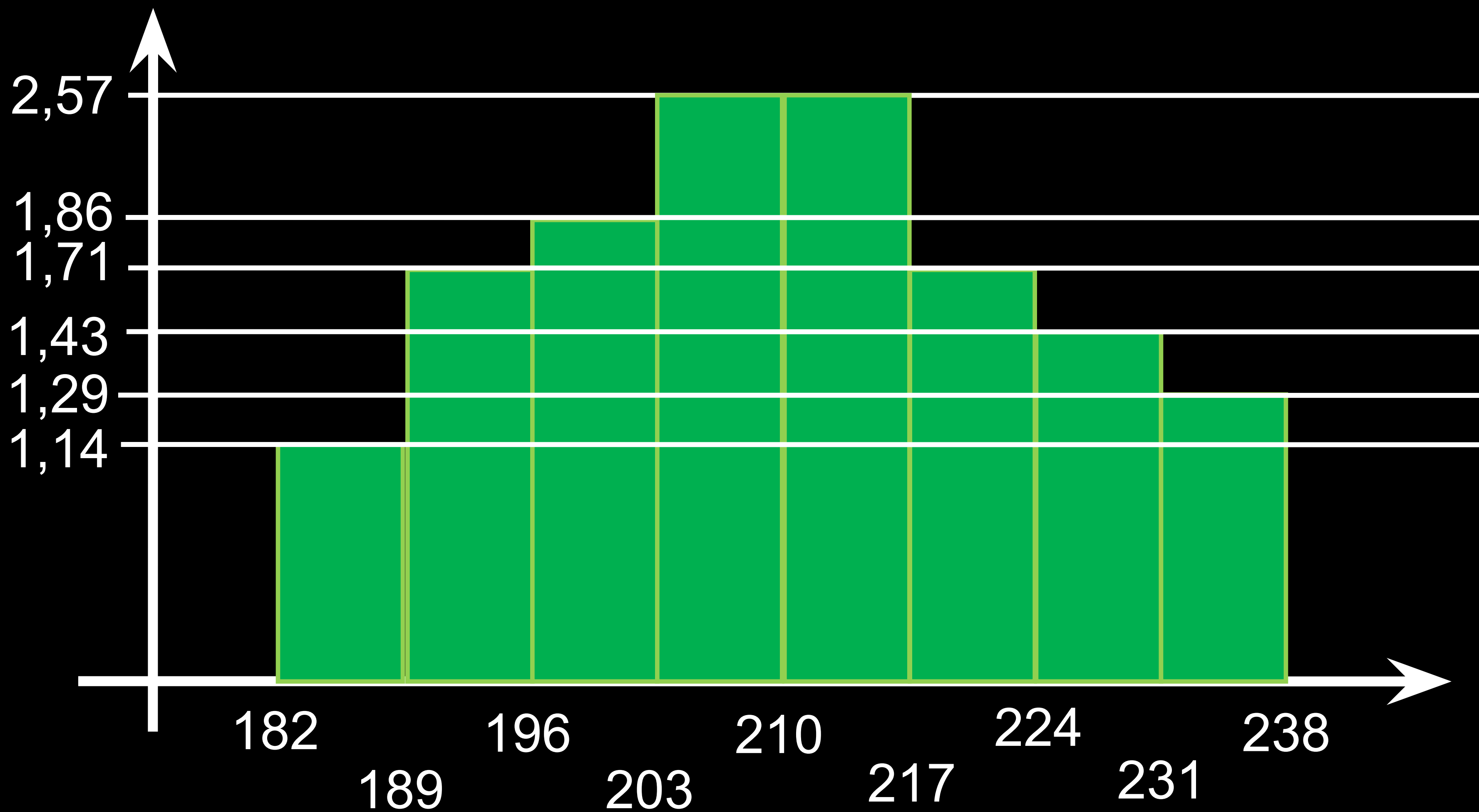
Замечание.

Площадь каждого прямоугольника:

Площадь всей гистограммы:

— объем выборки

Пример.



Гистограммой относительных частот называют ступенчатую фигуру, состоящую из прямоугольников с основанием h и высотой

<u>Замечание.</u>

Площадь каждого прямоугольника:

Площадь всей гистограммы:

Гистограмма относительных частот служит для оценки вида плотности вероятности.

п.4. Эмпирическая функция распределения.

Эмпирической функцией распределения называется функция, определяющая для каждого значения x относительную частоту события

где — число вариант, меньших x,

n — объем выборки.

Свойства эмпирической функции распределения

1)

2) — неубывающая функция.

3) Если — наименьшая варианта, то

Если — наименьшая варианта, то

Пример. Вариационный ряд

X	2	6	10
n	12	18	30

Построить эмпирическую функцию распределения.

Решение. Объем выборки

Если , то

Если , то

Если , то

Если , то

Таким образом,

п.5. Статистические оценки.

Пусть имеется некоторая выборка значений СВ, с теоретической функцией распределения

Однако, вид этой функции неизвестен.

Требуется найти (оценить) какой-либо параметр этого распределения (мат. ожидание, дисперсию и т.д.).

Пусть — точное значение этого параметра (неизвестное).

Пусть — статистическая оценка параметра

Пусть последовательно производятся выборки объема n.

Тогда можно рассматривать как СВ, принимающую значения

Для того, чтобы оценка давала хорошее приближение оцениваемому параметру она должна удовлетворять требованиям:

несмещенность;

эффективность;

состоятельность.

Оценка называется несмещенной, если ее мат. ожидание равно оцениваемому параметру

Оценка называется эффективной, если ее дисперсия минимальна:

Оценка называется состоятельной, если при большом объеме выборки ее значение приближается к истинному:

п.6. Числовые характеристики выборки.

Рассмотрим вариационный ряд

X			$\ldots$	
n			$\ldots$	

Размахом варьирования называется число

Выборочным средним называется среднее арифметическое значение вариант

Замечание.

Выборочное среднее является несмещенной состоятельной оценкой математического ожидания.

Выборочной дисперсией называется среднее значение квадратов отклонения вариант от выборочного среднего

Несложно получить, что

Замечание.

Выборочное среднее является смещенной оценкой теоретической дисперсии.
Можно показать, что

В качестве несмещенной оценки дисперсии используется исправленная выборочная дисперсия

Выборочным средним квадратическим отклонением называется квадратный корень из выборочной дисперсии

Исправленным выборочным средним квадратическим отклонением называется величина S (корень из).

Замечание.

Выборочное среднее и выборочная дисперсия обладают теми же свойствами, что и мат. ожидание и дисперсия дискретной СВ.

Начальным моментом r-го порядка называется среднее значение r-х степеней вариант

При этом

Центральным моментом r-го порядка называется среднее значение отклонений в степени r среднего

При этом

Модой Mo вариационного ряда называется варианта, имеющая наибольшую частоту.

Модой Me вариационного ряда называется варианта, которая делит ряд на две части, равные по числу вариант.
Асимметрией называется величина

Замечание.
Асимметрия характеризует меру симметричности эмпирической кривой распределения относительно среднего значения.
Для нормального распределения

Эксцессом называется величина

Замечание.

Эксцесс характеризует степень островершинности эмпирической кривой распределения по сравнению с нормальной кривой.

Для нормального распределения

Пример. Вариационный ряд

X	10-30	30-50	50-70	70-90	90-110	110-130
n	1	3	10	30	50	6

Найти числовые характеристики.

Решение.

Представим интервальный ряд в виде дискретного (в качестве вариант берем середины интервалов).

X	20	40	60	80	100	120
n	1	3	10	30	50	6

X	20	40	60	80	100	120
n	1	3	10	30	50	6

Объем выборки

Выборочное среднее

Выборочная дисперсия

Выборочное среднее квадратическое отклонение

Исправленная выборочная дисперсия

Исправленное выборочное среднее квадратическое отклонение

Мода

Медиана

Асимметрия

Отрицательная асимметрия говорит о том, что в вариационном ряде преобладают варианты, меньшие выборочного среднего.

Эксцесс

Искомая кривая распределения имеет более острую вершину по сравнению с нормальным распределением.