Лекция № 3.
Тема: «Математическая статистика»

Специальность: «Лечебное дело»
Курс: 2
Дисциплина: «Математика»

Подготовила: преподаватель высшей категории
Фёдорова Олеся Николаевна

Калуга 2010 год

В математической статистике разрабатываются теории и методы *обработки информации* о массовых явлениях и их назначении

Для этого проводится *статистическое исследование*, материалом для которого являются статистические данные

Статистические данные – это сведения о числе объектов какого - либо множества, обладающих некоторым признаком

Пример.

Сведения о числе отличников в каждом ССУЗе, сведения о числе разводов на число вступивших в брак

На основании статистических данных можно делать научно – обоснованные выводы

Для этого статистические данные определенным образом должны быть систематизированы и обработаны

Математическая статистика *изучает математические методы систематизации, обработки и использования* статистических данных для научных и производственных целей

Основной метод обработки данных – *выборочный*

Основа - *теория вероятности*, в которой изучаются математические модели реальных случайных явлений

Математическая статистика *связывает реальные случайные явления и их математические вероятностные модели*

Математическая статистика возникла в 17 веке одновременно с теорией вероятности

Статистическое исследование

Сплошное

Исследуется каждый объект совокупности

Выборочное

Исследуется отобранные некоторым образом объекты

Генеральная совокупность – совокупность всех исследуемых объектов

Выборочная совокупность (выборка) – совокупность случайно отобранных объектов

Случайный отбор – это такой отбор, при котором все объекты генеральной совокупности имеют одинаковую вероятность попасть в выборку

Выборка

повторная

бесповторная

Объект извлекается из генеральной совокупности, исследуется и возвращается в генеральную совокупность, берется следующий, исследуется и возвращается и т.д.

Объект извлекается из и не возвращается, берется генеральной совокупности, исследуется следующий

Объём выборки – это число равное количеству объектов генеральной или выборочной совокупности

Пример.

Из 10000 изделий для контроля отобрали 100 изделий

Объем генеральной совокупности равен 10000, объем выборки – 100

Математическая статистика занимается *вопросом*: можно ли установив *свойство выборки*, считать, что оно присуще *всей генеральной* совокупности

Для этого выборка должна быть достаточно ***представительной***, т.е. достаточно полно отражать изучаемое свойство объектов

Поэтому отбор объектов в выборку осуществляется *случайно*, а изучаемому свойству должна быть присуща ***статистическая устойчивость***: при многократном повторении исследования наблюдаемые события повторяются достаточно часто (статистическая устойчивость частот)

Для статистической обработки результаты исследования объектов, составляющих выборку, представляют в виде **числовой выборки** (последовательность чисел) $x_1, x_2, \ldots, x_n$

Разность между наибольшим значением числовой выборки и наименьшим называется **размахом выборки**

Рассмотрим числовую выборку объема n, полученную при исследовании некоторой генеральной совокупности

Значение x_1 встречается в выборке n_1 раз

x_2 встречается n_2 раза

…….

x_n встречается n_n раз

Числа $n_1, n_2, \ldots, n_n$ называются **частотами значений**

Отношения частот к объему выборки

$$\frac{n_1}{n}, \frac{n_2}{n}, \ldots, \frac{n_n}{n}$$

называются **относительными частотами значений**

$$n_1 + n_2 + \ldots + n_n = n \qquad \frac{n_1}{n} + \frac{n_2}{n} + \ldots + \frac{n_n}{n} = 1$$

Если составлена таблица в первой строке значения выборки, а во второй частоты значений, то она задает **статистический ряд**, если второй строке относительные частоты значений, то такая таблица задает **выборочное распределение**

x_1	x_2	x_3	…	x_n
n_1	n_2	n_3	…	n_n

x_1	x_2	x_3	…	x_n
n_1/n	n_2/n	n_3/n	…	n_n/n

Пример.

Для выборки определить объем, размах, найти статистический ряд и выборочное распределение:

3, 8, -1, 3, 0, 5, 3, -1, 3, 5

Объем: n = 10, размах = 8 – (-1) =9

Статистический ряд:

x_i	-1	0	3	5	8
n_i	2	1	4	2	1

Выборочное распределение:

x_i	-1	0	3	5	8
$\frac{n_i}{n}$	0,2	0,1	0,4	0,2	0,1

(убеждаемся 0,2 + 0,1 + 0,4 + 0,2 + 0,1 = 1)

Графические изображения выборки

Если выборка задана значениями и их частотами или статистическим рядом, то строится *полигон*

Полигон частот

Это ломаная с вершинами в точках

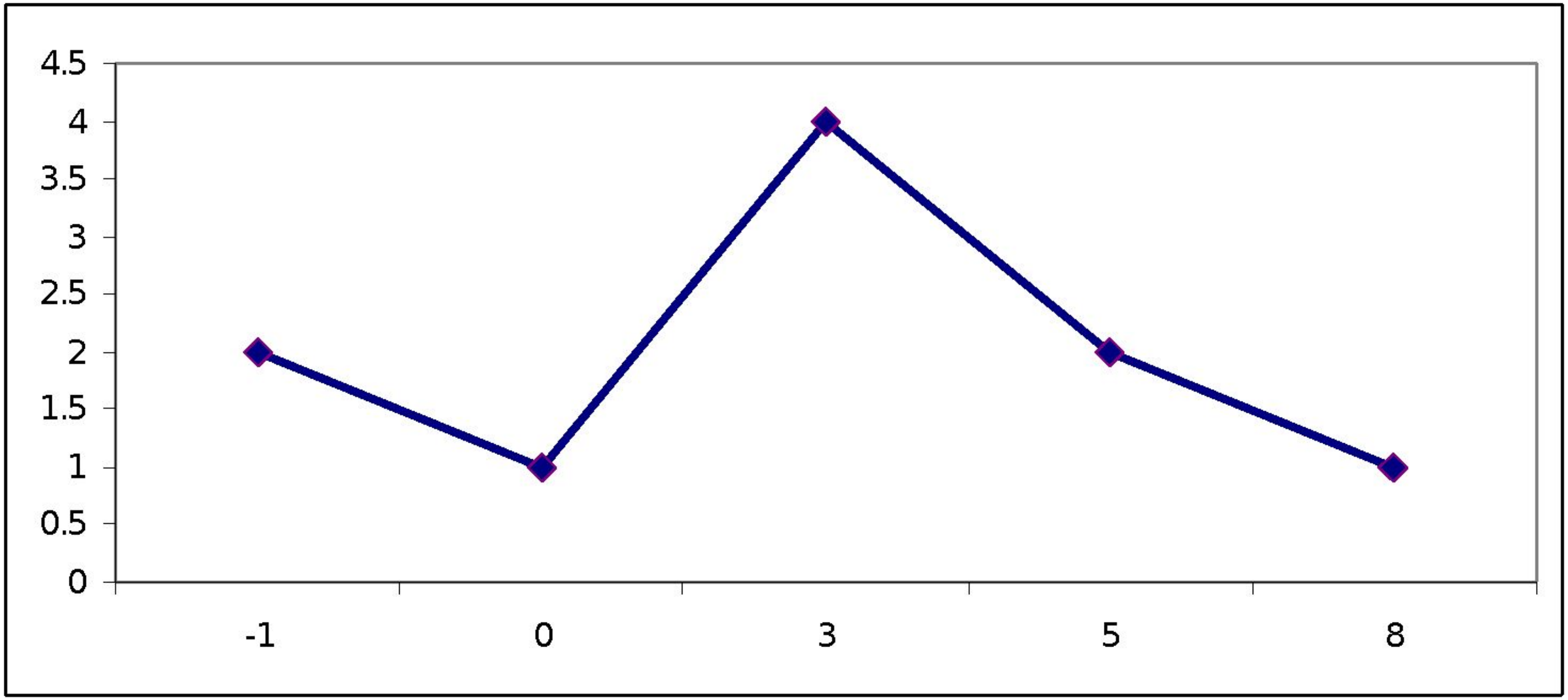
$$(x_1; n_1), (x_2; n_2), \ldots, (x_n; n_n)$$

Полигон относительных частот

Это ломаная с вершинами в точках

$$\left(x_1; \frac{n_1}{n}\right), \left(x_2; \frac{n_2}{n}\right), \ldots, \left(x_n; \frac{n_n}{n}\right)$$

Полигон частот



При большом объеме выборки строится *гистограмма*

Гистограмма частот **гистограмма относительных частот**

Для построения гистограммы промежуток от наименьшего значения выборки до наибольшего разбивают на несколько частичных промежутков длины h

Для каждого частичного промежутка подсчитывают сумму частот значений выборки, попавших в этот промежуток (S_i)

Значение выборки, совпавшее с правым концом частичного промежутка (кроме последнего промежутка), относится к следующему промежутку

Затем на каждом промежутке, как на основании, строим прямоугольник с высотой $\frac{S_i}{h}$

Ступенчатая фигура, состоящая из таких прямоугольников, называется **гистограммой частот**

Площадь такой фигуры равна ***объёму выборки***

Гистограммой относительных частот называют ступенчатую фигуру, состоящую из прямоугольников, основанием которых являются частичные промежутки длины h, а высотой отрезки длиной

$$\frac{\omega_i}{h}$$

где ω_i – сумма относительных частот значений выборки, попавших в i промежуток

Площадь такой фигуры ***равна 1***

Пример.

В результате измерения напряжения в электросети получена выборка. Построить гистограмму частот, если число частичных промежутков равно 5

218, 224, 222, 223, 221, 220, 227, 216, 215, 220, 218, 224, 225, 219, 220, 227, 225, 221, 223, 220, 217, 219, 230, 222

$n = 24$

Наибольшее значение – 230

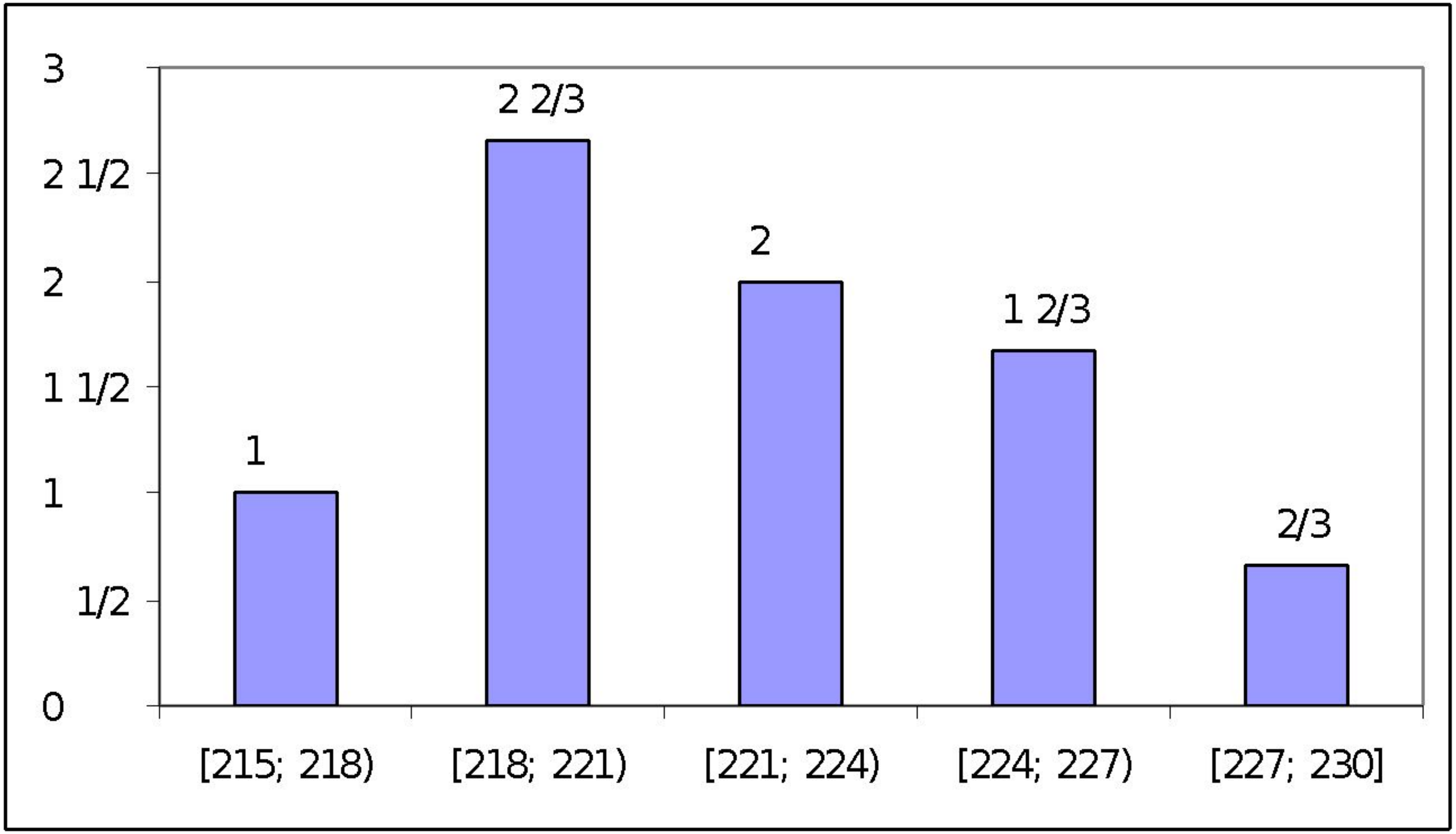
Наименьшее значение – 215

Интервал: 230 – 215 = 15

Длина частичных промежутков: $h = \frac{15}{5} = 3$

Составим таблицу:

№	интервал	S_i	$\frac{S_i}{h}$
1	[215; 218)	3	$\frac{3}{3}=1$
2	[218; 221)	8	$\frac{8}{3}=2\frac{2}{3}$
3	[221; 224)	6	$\frac{6}{3}=2$
4	[224; 227)	4	$\frac{4}{3}=1\frac{1}{3}$
5	[227; 230]	3	$\frac{3}{3}=1$


3
2 1/2
2
1 1/2
1
1/2
0
1
2 2/3
2
1 2/3
2/3
[215; 218)
[218; 221)
[221; 224)
[224; 227)
[227; 230]

Выборочные характеристики

Для выборки объема n $x_1, x_2, ..., x_n$

Выборочное статистическое ожидание (выборочное среднее) – это среднее арифметическое значений выборки

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$

Если выборка задана статистическим рядом, то

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + ... + n_n x_n}{n}$$

Выборочная дисперсия – это среднее арифметическое квадратов отклонений значений выборки от выборочного среднего

$$S_0 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n}$$

Если выборка задана статистическим рядом, то

$$S_0 = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \ldots + n_n(x_n - \bar{x})^2}{n}$$

Несмещенная выборочная дисперсия

$$S = \frac{n}{n-1} \cdot S_0$$

Пример.

Для выборки найти $\bar{x}, S_0, S$

Выборка: 4, 5, 3, 2, 1, 2, 0, 7, 7, 3

n = 10

$$\bar{x} = \frac{4+5+3+2+1+2+0+7+7+3}{10} = \frac{34}{10} = 3{,}4$$

$$S_0 = \frac{(4-3{,}4)^2 + (5-3{,}4)^2 + (3-3{,}4)^2 + (2-3{,}4)^2 + (1-3{,}4)^2 +}{10}$$

$$\frac{+(2-3{,}4)^2 + (0-3{,}4)^2 + (7-3{,}4)^2 + (7-3{,}4)^2 + (3-3{,}4)^2}{10} = \frac{50{,}4}{10} = 5{,}04$$

$$S = \frac{10}{9} \cdot 5{,}04 = \frac{50{,}4}{9} = 5{,}6$$