Предмет математической статистики.

Генеральная и выборочная совокупности.

Основные вопросы:

- ***Основные задачи математической статистики.***
- ***Основные понятия математической статистики: генеральная и выборочная совокупности.***

Определение

Математическая статистика – это раздел математики, который изучает методы обработки и классификации статистических данных для получения научно-обоснованных выводов и принятия решений.

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

Математическая статистика возникла в XVII веке и развивалась параллельно с теорией вероятностей. Дальнейшее развитие (вторая половина XIX века – начало XX века) обязано, в первую очередь, П. Л. Чебышеву, А. А. Маркову, А. М. Ляпунову, а так же К. Гауссу, А. Кетле, Ф. Гальтону, К. Пирсону и другие. XX век – советские учёные : В. И. Романовский, Е. Е. Слуцкий, А. Н. Колмогоров. Английские: Стьюдент, Фишер, Смирнов. Американские:С. Нейман, Вальд.

Задачи математической статистики

1. ***Оценка неизвестных параметров*** случайной величины (вероятности случайного события, математического ожидания случайной величины, дисперсии)
2. ***Статистическая проверка гипотез***, т.е. проверка предположений, сделанных относительно некоторых случайных событий, случайных величин (о вероятности события, о законе распределения случайной величины)
3. ***Принятие решений*** (сюда относятся задачи оптимального выбора момента настройки или замены действующей аппаратуры, например, определения срока замены двигателя самолета, отдельных деталей станков)

Генеральная и выборочная совокупность

Пусть требуется изучить совокупность однородных объектов относительно некоторого ***качественного*** или ***количественного*** признака, характеризующего эти объекты.

Определения выборочной и генеральной совокупности

Выборочной совокупностью **или** ***выборкой*** называют совокупность случайно отобранных объектов.

Генеральной совокупностью называют совокупность объектов из которых производится выборка.

Объемом совокупности называют число объектов этой совокупности. Например, если из 1000 деталей отбирается для обследования 100, то *объем генеральной совокупности* $N=1000$, а *объем выборки* $n = 100$.

- При составлении выборки можно поступать **двумя способами**: после того как объект отобран и исследован, его можно возвратить или не возвращать в генеральную совокупность.
- В связи с этим выборки подразделяются на **повторные** и **бесповторные**.
- ***Повторной*** называют выборку, при которой отобранный объект (перед отбором следующего) ***возвращается*** в генеральную совокупность.
- При ***бесповторной*** выборке отобранный объект в генеральную совокупность ***не возвращается***.

Репрезентативность выборки.

Для того чтобы по данным выборки можно было достаточно уверенно судить об интересующем признаке *генеральной совокупности*, необходимо, чтобы объекты выборки правильно его представляли. Другими словами, выборка должна правильно представлять пропорции генеральной совокупности. Это требование коротко формулирует так: выборка должна быть ***репрезентативной (представительной).***

В *силу закона больших чисел* можно утверждать, что выборка будет репрезентативной:

- каждый объект выборки ***отобран случайно*** из генеральной совокупности;
- все объекты имеют ***одинаковую вероятность*** попасть в выборку.

Способ отбора

- Простой
- Типический
- Механический
- Серийный

Способы отбора объектов наблюдения

Простой случайный отбор

Объект извлекают по одному из Генеральной совокупности с помощью генератора случайных чисел

- ***Бесповторный*** – исключать из рассмотрения объекты, которые уже попали в статистическую выборку
- ***Повторный*** – допускать возможность повторения объектов в статистической выборке

Способы отбора объектов наблюдения

Типический отбор

Объекты отбирают из каждой «**типической**» части генеральной совокупности.

Используется, если обследуемый признак заметно колеблется в различных частях генеральной совокупности.

Пример:

Продукция изготавливается на нескольких машинах с различной степенью изношенности. *Тогда отбор следует производить из продукции, выпущенной машинами определенного типа*

Способы отбора объектов наблюдения

Механический отбор

Генеральную совокупность «**механически**» делят на группы, их количество равно объему выборки, затем из каждой группы отбирают по одному объекту наблюдения.

Пример:

Если необходимо выбрать 20% изготавливаемых деталей, то отбирают каждую 5-ю деталь, если 5% деталей, то отбирают каждую 20-ю деталь

Способы отбора объектов наблюдения

<u>Серийный отбор</u>

Объекты отбирают «**сериями**», которые обследуются полностью.

Используется, когда обследуемый признак колеблется незначительно между сериями.

Пример: Если все детали производятся на одинаковых станках-автоматах, то достаточно выбрать несколько станков для сплошного обследования произведенных деталей.

<u>Комбинированный отбор</u>

Часто используется сочетание нескольких способов отбора объектов наблюдения: **генеральная совокупность разделяется на серии, серии на группы, из групп отбираются объекты**.

Пример . Десять абитуриентов проходят тестирование по математике. Каждый из них может набрать от 0 до 5 баллов включительно. Пусть X_k — количество баллов, набранных k-м ($k = 1, 2, \ldots, 10$) абитуриентом.

Тогда значения 0, 1, 2, 3, 4, 5 — все возможные количества баллов, набранных одним абитуриентом, — образуют генеральную совокупность.

Выборка $X_1, X_2, X_3, \ldots, X_{10}$ — результат тестирования 10 абитуриентов.

Реализациями выборки могут быть следующие наборы чисел: {5, 3, 0, 1, 4, 2, 5, 4, 1, 5} или {4, 4, 5, 3, 3, 1, 5, 5, 2, 5} или {3, 4, 5, 0, 1, 2, 3, 4, 5, 4} и т. д.

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

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 $n_1, n_2, \dots, n_n$ называются ***частотами значений***

Отношения частот к объему выборки

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

называются ***относительными частотами значений***

$$\boxed{n_1 + n_2 + \dots + n_n = n}$$

$$\boxed{\frac{n_1}{n} + \frac{n_2}{n} + \dots + \frac{n_n}{n} = 1}$$

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

Полигон относительных частот

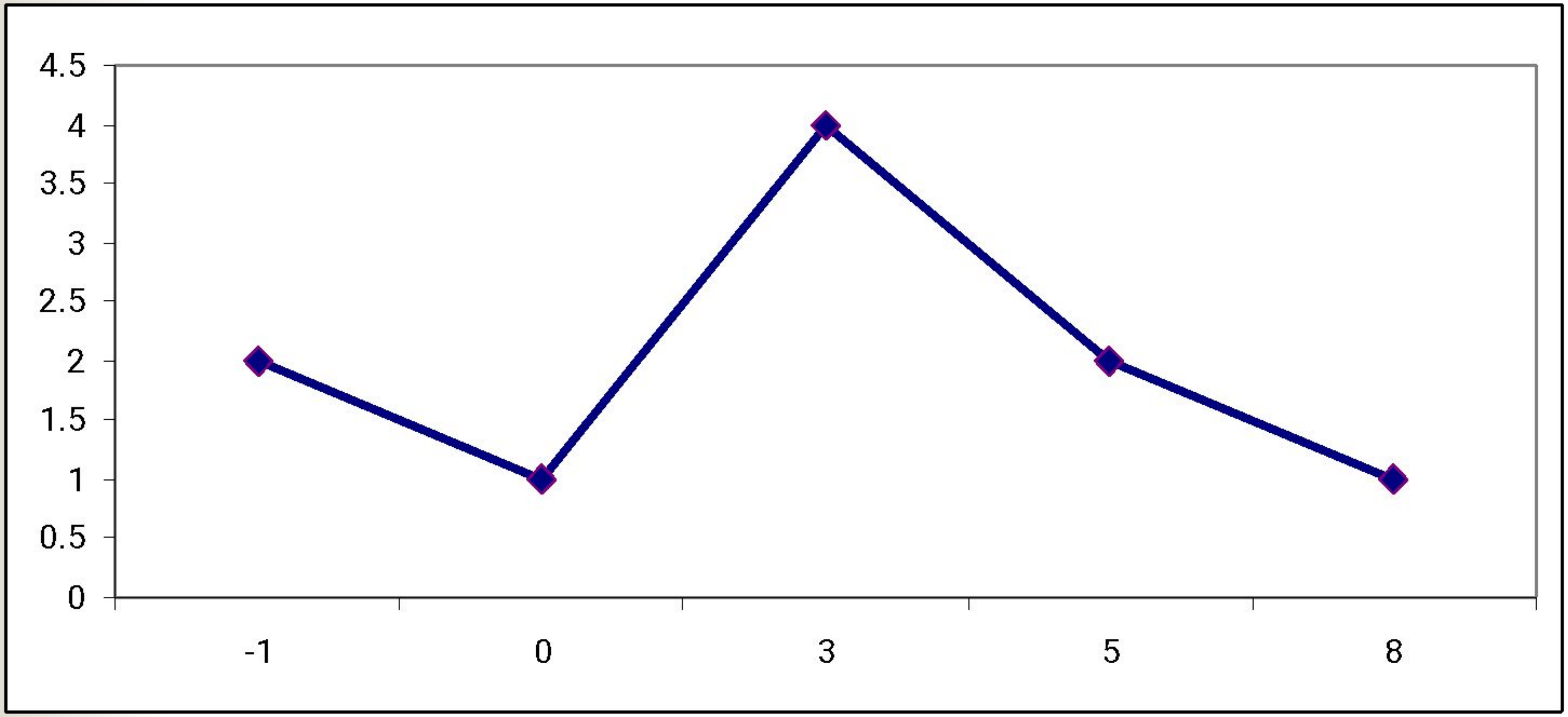
Это ломаная с вершинами в точках

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

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

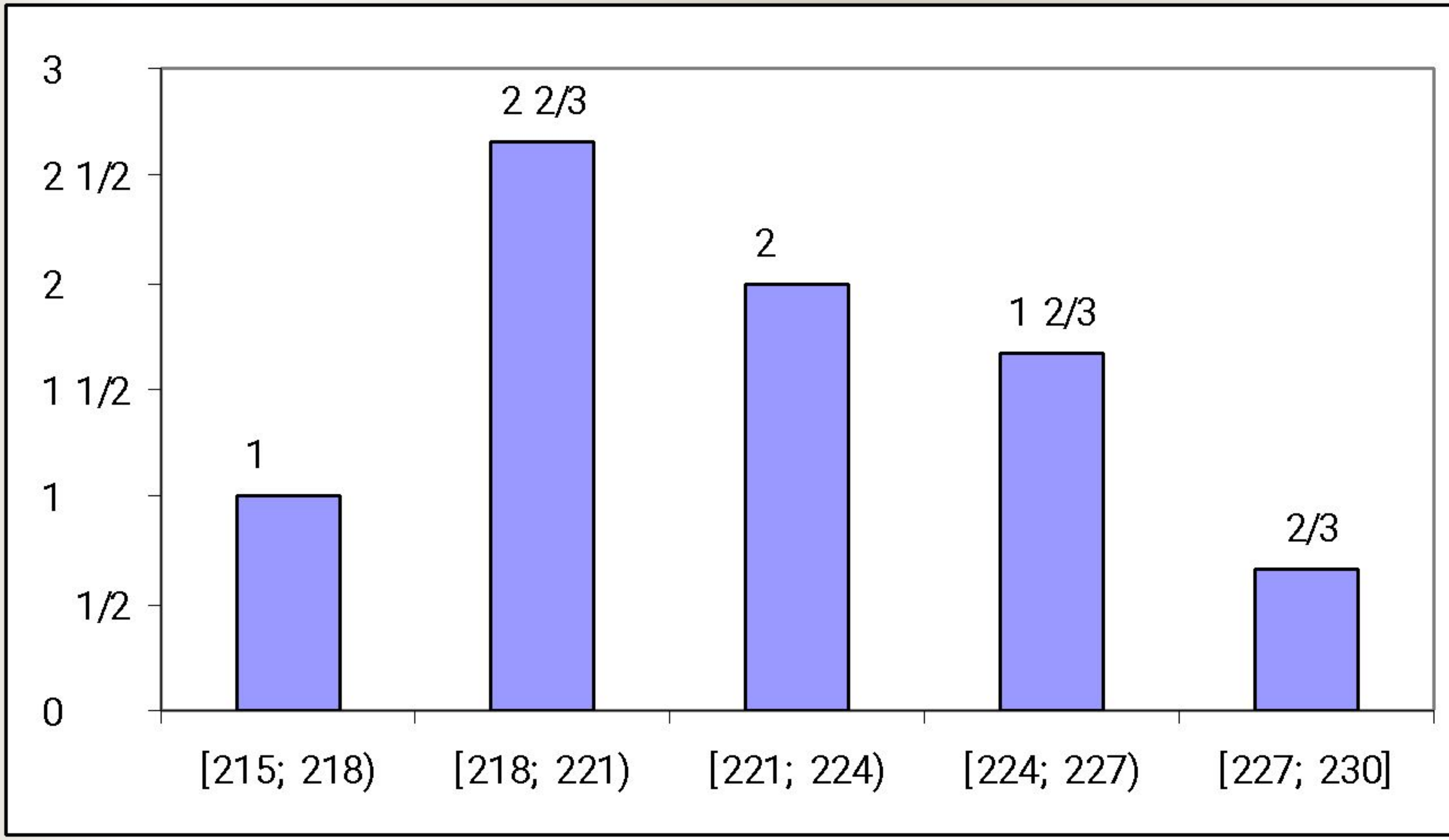
Наименьшее значение – 215

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

Выборочные характеристики

Для выборки объема **n** $x_1, x_2, \ldots, x_n$

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

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

Если выборка задана ***статистическим рядом***, то

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

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

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

Если выборка задана статистическим рядом, то

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

Несмещенная выборочная дисперсия

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

где S_0—выборочная дисперсия, n—объем выборки.

Пример.

Для выборки найти $\bar{x}, S_0, S$

Выборка: 4, 5, 3, 2, 1, 2, 0, 7, 7, 3

n = 10

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

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

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

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