ЛЕКЦИЯ
ТЕМА: «МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»

Термин статистика происходит от латинского слова **«status».**

В Средние века это означало политическое состояние государства.

В науку этот термин ввёл немецкий учёный **Годфрид Ахенваль.**

Зарождение статистики, как науки, следует отнести ко второй половине XVII века.

В настоящее время термин **«Статистика»** употребляется в четырёх значениях:

1. Комплекс дисциплин, обладающих определённой спецификой и изучающих количественную сторону массовых явлений и процессов в их неразрывной связи с их качественным содержанием – учебный предмет в ВУЗ-ах и СУЗ-ах;

2. **Отрасль практической деятельности** по сбору, обработке, анализу и публикации массовых цифровых данных о самых различных явлениях и процессах общественной жизни;

3. **Совокупность цифровых сведений**, характеризующих состояние массовых явлений и процессов общественной жизни;

4. Статистические методы, применяемые для изучения социально-экономических явлений и процессов.

Статистика как наука имеет свой предмет исследования. Она исследует не отдельные факты, а массовые социально-экономические явления и процессы, выступающие как множество отдельных факторов, обладающих как индивидуальными, так и общими признаками.

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

Пример.

Сведения о количестве отличников в каждом учебном заведении;

сведения о числе разводов на число вступивших в брак;

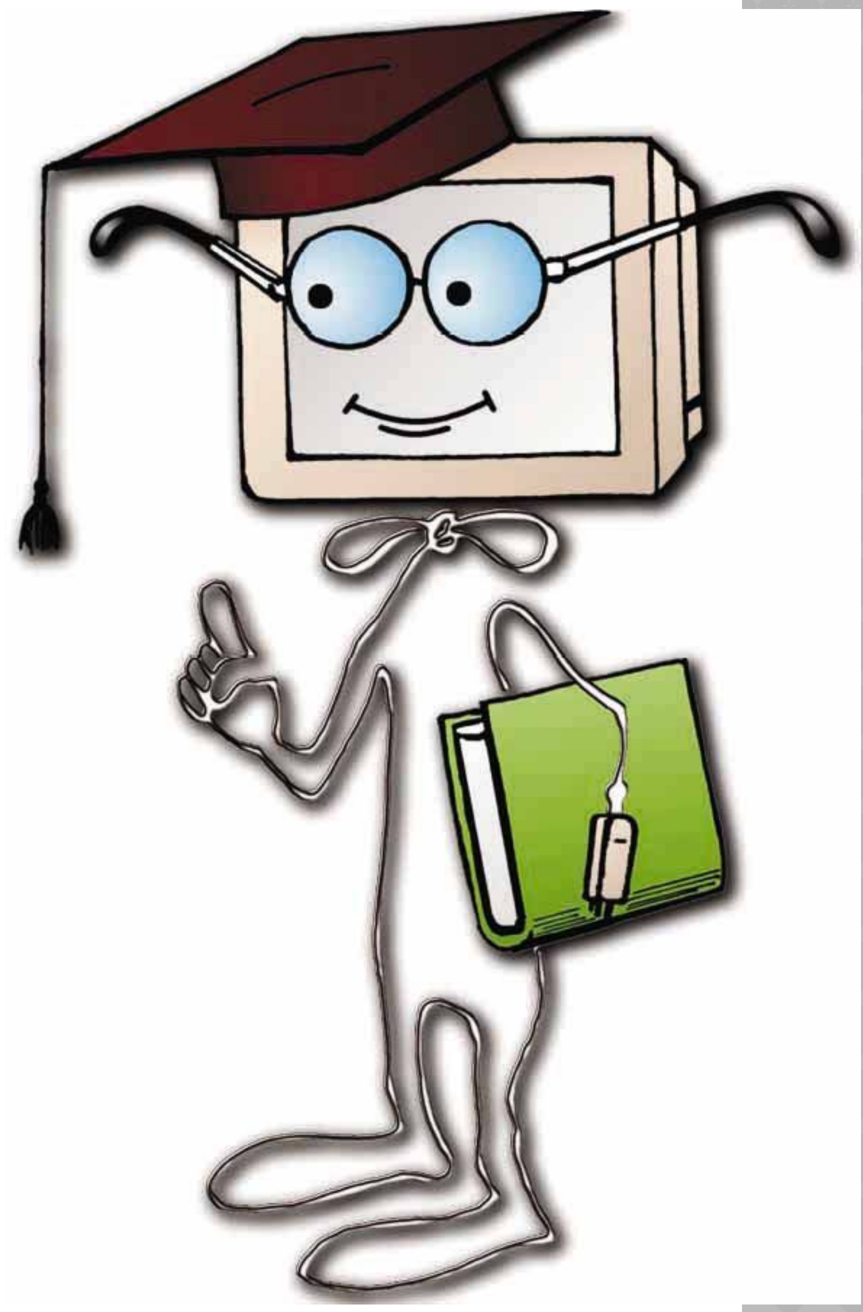
сведения о количестве новорожденных и др.

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 нужд.

Основной метод обработки данных – ***выборочный***

Явления и процессы в жизни общества изучаются статистикой посредством ***статистических показателей***, представляющих собой обобщённую числовую характеристику какого-либо явления в единстве с качественной стороной в условиях конкретного места и времени.



Различают следующие статистические показатели:

учётно-оценочные, которые в зависимости от специфики изучаемого явления могут отображать или объёмы их распространенности в пространстве или достигнутые на определённые моменты (даты) уровни развития. (Например: численность населения в России на начало 2002 года составила 146,3 млн. чел.);

аналитические показатели, применяются для анализа статистической информации и характеризуют особенности развития изучаемого явления: типичность признака, соотношение его отдельных частей, меру распространения в пространстве, скорость развития во времени и т.д. В качестве аналитических показателей в статистике применяются относительные и средние величины, показатели вариации и динамики, тесноты связи и др.

Различают следующие статистические показатели:

Одной из важных категорий статистической науки, тесно связанной с показателями, является понятие ***признака,*** под которым понимается характерное свойство изучаемого явления, отличающее его от других явлений.

Признаки бывают:

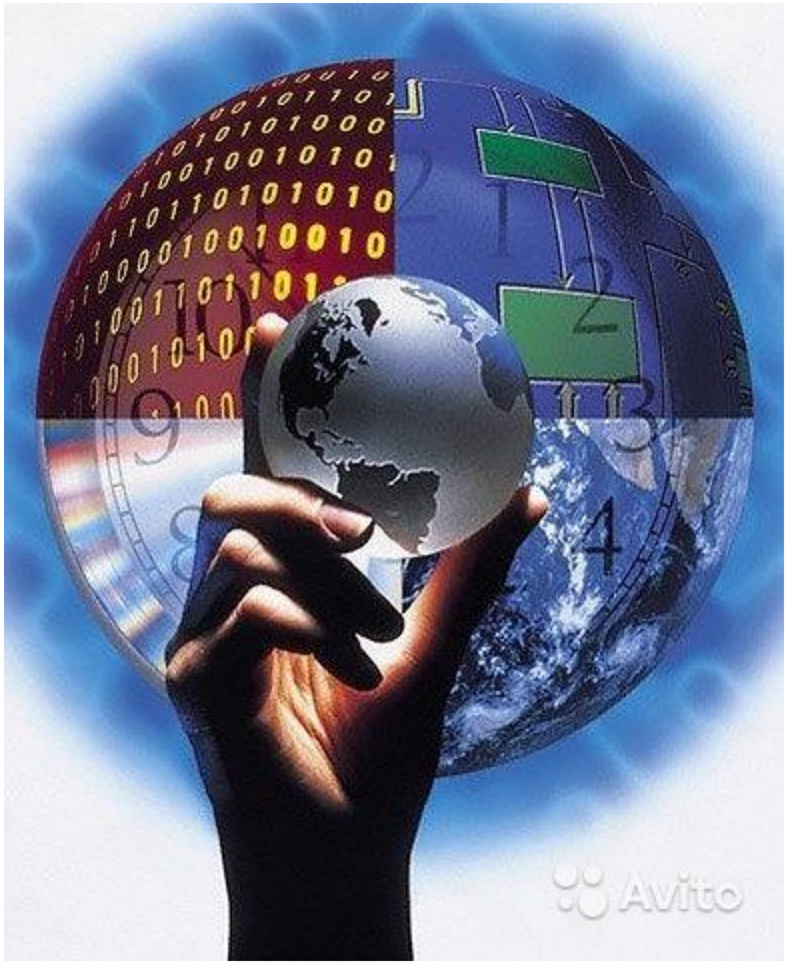
атрибутивные, выраженные смысловыми понятиями (пол – мужской, женский; магазин – продовольственный, промтоварный, хозяйственный);

количественные – признаки, выраженные числовыми значениями (возраст человека, стаж работы, размер заработной платы и т.д.);

варьирующие, принимающие различные значения у отдельных единиц изучаемого явления (товарооборот, валовой сбор и т.д.).

Статистика рассматривает статистические совокупности.

Статистическая совокупность представляет собой множество единиц изучаемого явления, объединённых в соответствии с задачей исследования единой качественной стороной, т. е. ***признаками***



Целью изучения статистических совокупностей является выявление закономерностей. ***Закономерность*** – это то общее что определяет единство и однородность совокупности.

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

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

бесповторная

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 отража   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** (последовательности чисел или числового ряда)

$$x_1, x_2, \ldots, x_n$$

Показатели описательной статистики можно разбить на несколько групп:

- показатели положения, описывающие положение экспериментальных данных на числовой оси. Примеры таких данных – ***максимальный и минимальный элементы выборки***, ***среднее значение***, ***медиана***, ***мода*** и др.;

- показатели разброса, описывающие степень разброса данных относительно центральной тенденции. К ним относятся: выборочная дисперсия, разность между минимальным и максимальным элементами (***размах, интервал выборки***) и др.;

- показатели асимметрии: ***положение медианы относительно среднего*** и др.;

- графические представления результатов – ***гистограмма, частотная диаграмма*** и др.

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**

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

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

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)

Мода (**m_o**) — это наиболее частое значение в выборке, или среднее значение класса с наибольшей частотой. Мода как центральная тенденция используется чаще всего для того, чтобы дать общее представление о распределении. В некоторых случаях у распределения могут быть две моды, в таком случае это свидетельствует о бимодальном распределении, что указывает на наличие двух относительно самостоятельных групп.

Медиана (**m_e**) соответствует центральному значению в последовательном ряду всех полученных значений выстроенном в порядке возрастания. Если же в ряду чётное количество показателей , то берут среднее арифметическое двух средних значений

Среднее арифметическое (**m**) — это показатель центральной тенденции, полученный делением суммы всех значений данных на число этих данных. Среднее арифметическое используется для представления количественных переменных с нормальным распределением.

Указание в представлении данных меры центральной тенденции (среднее, медиана, мода) автоматически сообщает ***о нормальности распределения признака***.

При нормальном распределении все три показателя более или менее ***совпадают***, а при асимметричном распределении — ***нет***.

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

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

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

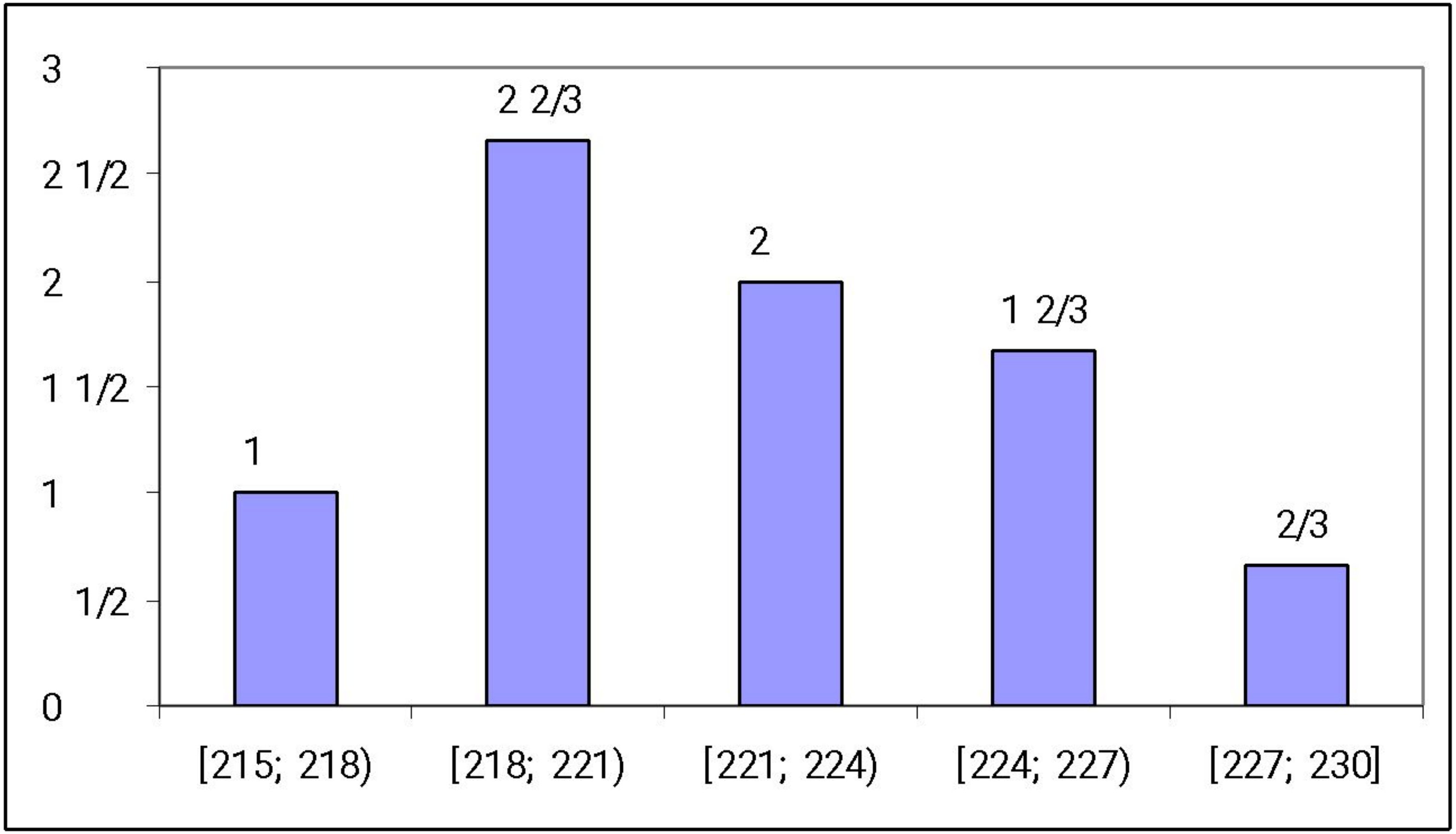
Наименьшее значение – 215

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 ожидание (выборочное среднее) – это среднее арифметическое значений выборки

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

Если выборка задана статистическим рядом, то

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

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

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

Если выборка задана статистическим рядом, то

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

Несмещенная выборочная дисперсия

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

Пример.

Для выборки найти $\bar{x}, S_0, S$

Выборка: 4, 5, 3, 2, 1, 2, 0, 7, 7, 3

n = 10

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

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

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

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