Лекция 8

Тема: Статистическое распределение выборки и его основные числовые характеристики

- Предмет математической статистики
- Основные числовые характеристики выборки
- Графическое представление выборки

Предмет математической статистики

Математическая статистика базируется на теории вероятностей и является теоретической основой всей статистики. Ее задачей является создание способов сбора и методов обработки статистической информации.

Выборочный метод – один из основных методов математической статистики. Его сущность заключается в том, что изучение большой совокупности объектов относительно некоторого количественного признака X производится по сравнительно небольшому числу случайно отобранных объектов.

Генеральной совокупностью называется множество всех изучаемых объектов, из которых производится выборка.

Выборочной совокупностью (выборкой) называется множество всех изучаемых объектов, отобранных для изучения из генеральной совокупности.

Выборка должна быть организована случайным образом, чтобы правильно представлять генеральную совокупность.

Объемом совокупности называется количество объектов в совокупности. Объем выборки n, как правило, значительно меньше объема N генеральной совокупности: $n<<N$.

Данные выборки записываются в виде таблицы, называемой **статистическим распределением выборки:**

x_i	x_1	x_2	…	…	x_k
n_i	n_1	n_2	…	…	n_k

X_i – варианты, располагаемые в порядке возрастания или убывания; n_i – частоты, показывающие сколько раз значение варианты наблюдалось в выборке

Основные числовые характеристики выборки

1.Выборочная средняя – это значение, которым можно, в среднем, заменить все значения выборки (среднее взвешенное значение признака в выборке):

$$\bar{x}_в = \frac{1}{n}\left(x_1 n_1 + x_2 n_2 + \ldots + x_k n_k\right)$$

2.Выборочная дисперсия – характеризует разброс (рассеяние) значений вариант x_i от выборочного среднего и измеряется в квадратных единицах признака X:

$$D_в = \frac{1}{n}(x_1^2 n_1 + x_2^2 n_2 + \ldots + x_k^2 n_k) - (\bar{x}_в)^2$$

3.Среднее квадратическое отклонение – характеристика рассеяния значений признака в выборке от среднего выборочного в единицах признака *X*:

$$\sigma_в = \sqrt{D_в}$$

С помощью найденных выборочных характеристик можно оценить соответствующие характеристики генеральной совокупности (генеральные характеристики).

$$\bar{x}_в, D_в, \sigma_в$$

Оценки имеют следующий вид:

$$\bar{x} \approx \bar{x}_{в} \qquad D \approx \frac{n}{n-1} \cdot D_{в} \qquad \sigma \approx \sqrt{D}$$

Приведенные оценки носят случайный характер, так как зависят от выборки. Они называются **точечными** и удовлетворяют следующим требованиям:

- *несмещенность* (отсутствие систематических ошибок);
- *состоятельность* (увеличение объема выборки повышает вероятность правильности оценки);
- *эффективность* (имеют самый незначительный разброс по сравнению с другими возможными оценками).

Статистическое представление выборки

Вид статистического распределения зависит от объема выборки.

Если объем выборки невелик ($n<30$), то строят **дискретное статистическое распределение.**

Если объем выборки $n>>30$, то строят **интервальное статистическое распределение.**

Пример построения дискретного статистического распределения

Пример 1 Известны следующие данные о результатах сдачи абитуриентами вступительных экзаменов в ВУЗ (баллов):

18,16,20,17,19,20,17,17,12,15,20,18,19,18,18,16,18,14,14,17,19 ,16,14,19,12,15,16,20. (n=28)

Построим ряд распределения абитуриентов, распределив их по количеству набранных баллов:

12(2), 14(3), 15(2), 16(4), 17(4), 18(5), 19(4), 20(4),

где скобках указываем частоты появления конкретного количества набранных баллов

Те же данные можно представить в виде таблицы распределения:

x_i	12	14	15	16	17	18	19	20
n_i	2	3	2	4	4	5	4	4

Вычислим числовые характеристики данного распределения:

Пример построения интервального статистического распределения

Пример 2

Распределение населения РФ в 1 квартале 1996 г.

по уровню среднедушевых денежных доходов (N=100 чел.)

Среднедушевой денежный доход в среднем за месяц, тыс.руб.	200- 400	400-600	600-800	800-1000	1000-1200	1200-1400	1400-1600
Численность населения	33	24	16	10	7	6	4

Для вычисления числовых характеристик и графического представления выборки в полученную таблицу добавляют две строки:

Среденедушевой денежный доход в среднем за месяц, тыс.руб.	200-400	400-600	600-800	800-1 000	1000-1200	1200-1400	1400-1600
Численность населения (n_i)	33	24	16	10	7	6	4
y_i	300	500	700	900	1100	1300	1500
$\frac{n_i}{N}$	0,33	0,24	0,16	0,1	0,07	0,06	0,04

y_i – середина интервала; $\frac{n_i}{N}$- относительная частота

Вычислим числовые характеристики интервального ряда распределения (выборки):

$$\bar{x}_в = \frac{1}{100}\left(300\cdot 33 + 500\cdot 24 + 700\cdot 16 + 900\cdot 10 + 1100\cdot 7 + 1300\cdot 6 + 1500\cdot 4\right) \approx 636$$

$$D_в = \frac{1}{100}\left(300^2\cdot 33 + 500^2\cdot 24 + 700^2\cdot 16 + 900^2\cdot 10 + 1100^2\cdot 7 + 1300^2\cdot 6 + 1500^2\cdot 4\right) - 636^2 \approx 120704$$

$$\sigma = \sqrt{120704} \approx 347{,}4$$

По найденным числовым характеристикам выборки оценим числовые характеристики генеральной совокупности:

$$\bar{x} \approx \bar{x}_в \approx 636$$

$$D \approx \frac{100}{100-1} D_в \approx \frac{100}{99}\cdot 120704 \approx 121923{,}2$$

$$\sigma \approx \sqrt{121923{,}2} \approx 349{,}2$$

Таким образом, для данной выборки среднедушевой доход в первом квартале 1996 года составлял 636 тыс.рублей; отклонение доходов членов выборки от среднедушевого дохода составляло 349,2 тыс. рублей.

Графическое представление выборки

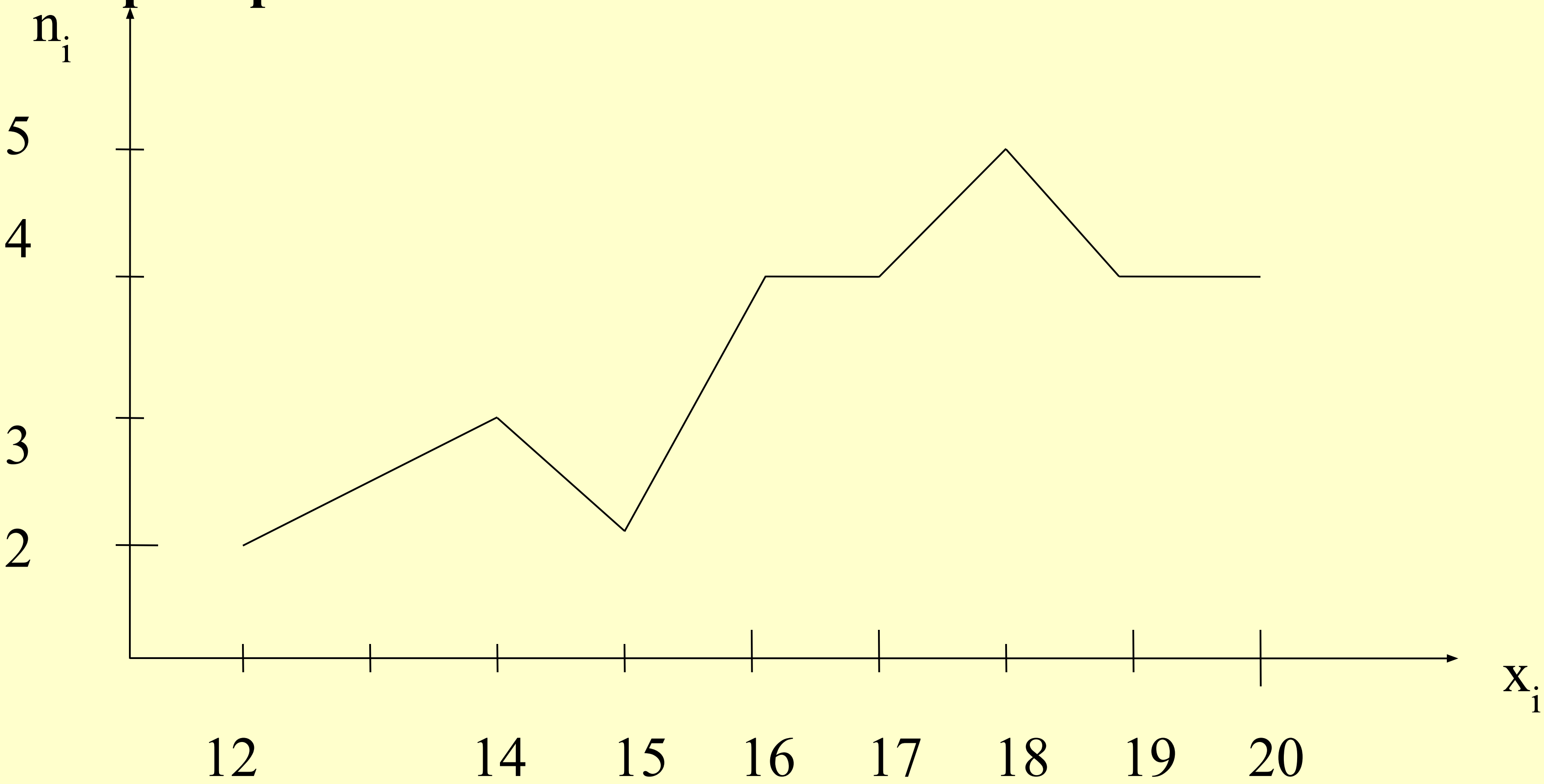
Графически дискретное статистическое распределение изображается в виде **полигона частот**, обычно относительных. **Полигон** представляет собой ломаную линию, соединяющую соседние точки с координатами (x_i, n_i).

Интервальное статистическое распределение изображается на графике в виде **гистограммы относительных частот**.

Гистограмма – это ступенчатая фигура, состоящая из прямоугольников. В основании каждого прямоугольника лежит частичный интервал, а высота прямоугольника равна относительной частоте .

При таком построении сумма площадей всех прямоугольников, входящих в гистограмму будет равна 1.

Построим полигон для дискретного распределения из **примера 1**:

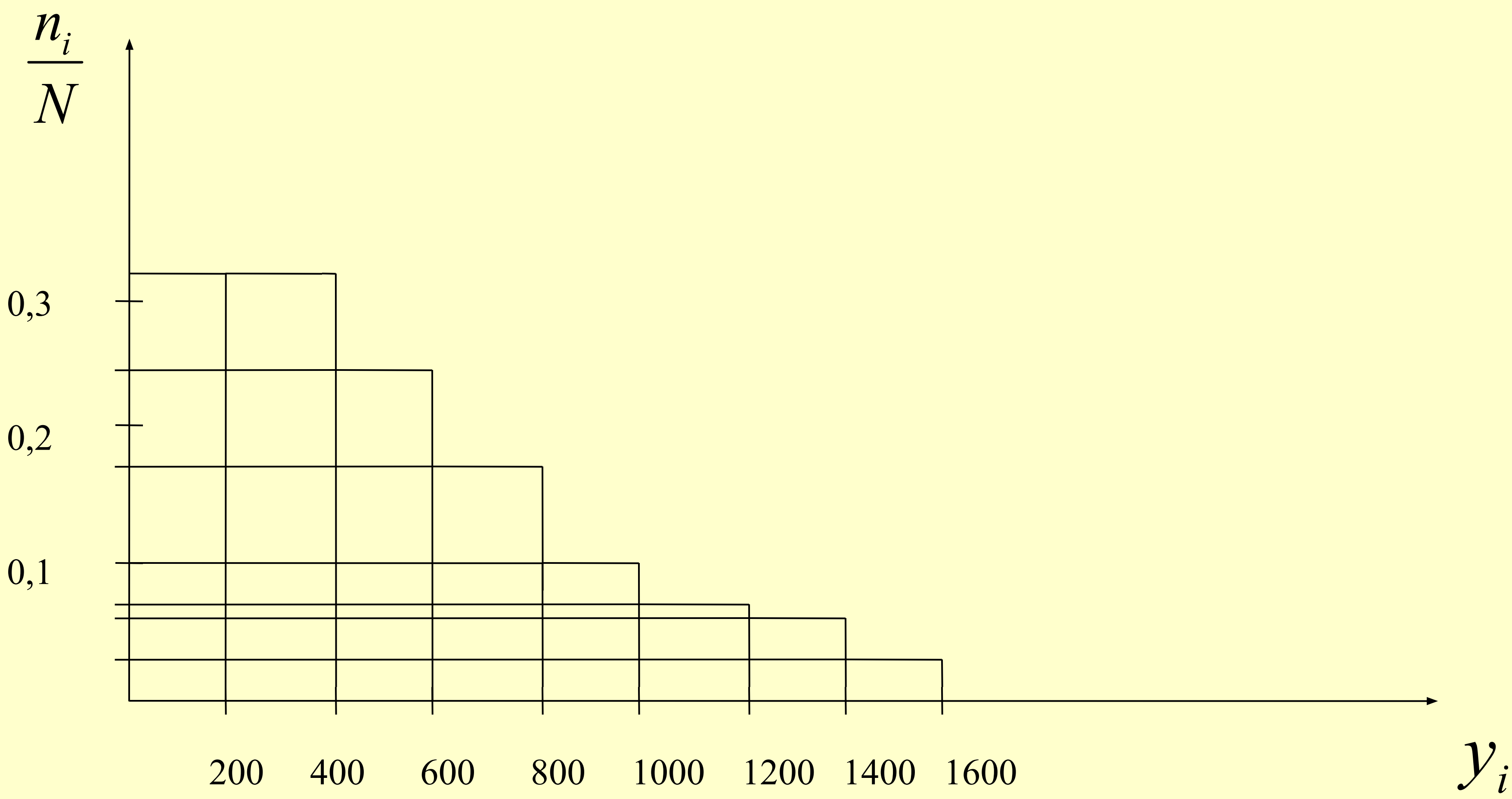


Интервальное статистическое распределение изображается на графике в виде **гистограммы относительных частот**.

Гистограмма – это ступенчатая фигура, состоящая из прямоугольников. В основании каждого прямоугольника лежит частичный интервал, а высота прямоугольника равна относительной частоте .

При таком построении сумма площадей всех прямоугольников, входящих в гистограмму, будет равна 1.

Построим гистограмму для интервального распределения из **примера 2:**



Вопросы для самопроверки

- В чем заключается суть предмета математической статистики?
- Что такое выборка? Что такое генеральная совокупность?
- Какими свойствами должна обладать выборка?
- Назовите числовые характеристики выборки.
- Как они связаны с числовыми характеристиками генеральной совокупности?
- Назовите способы графического представления выборки?