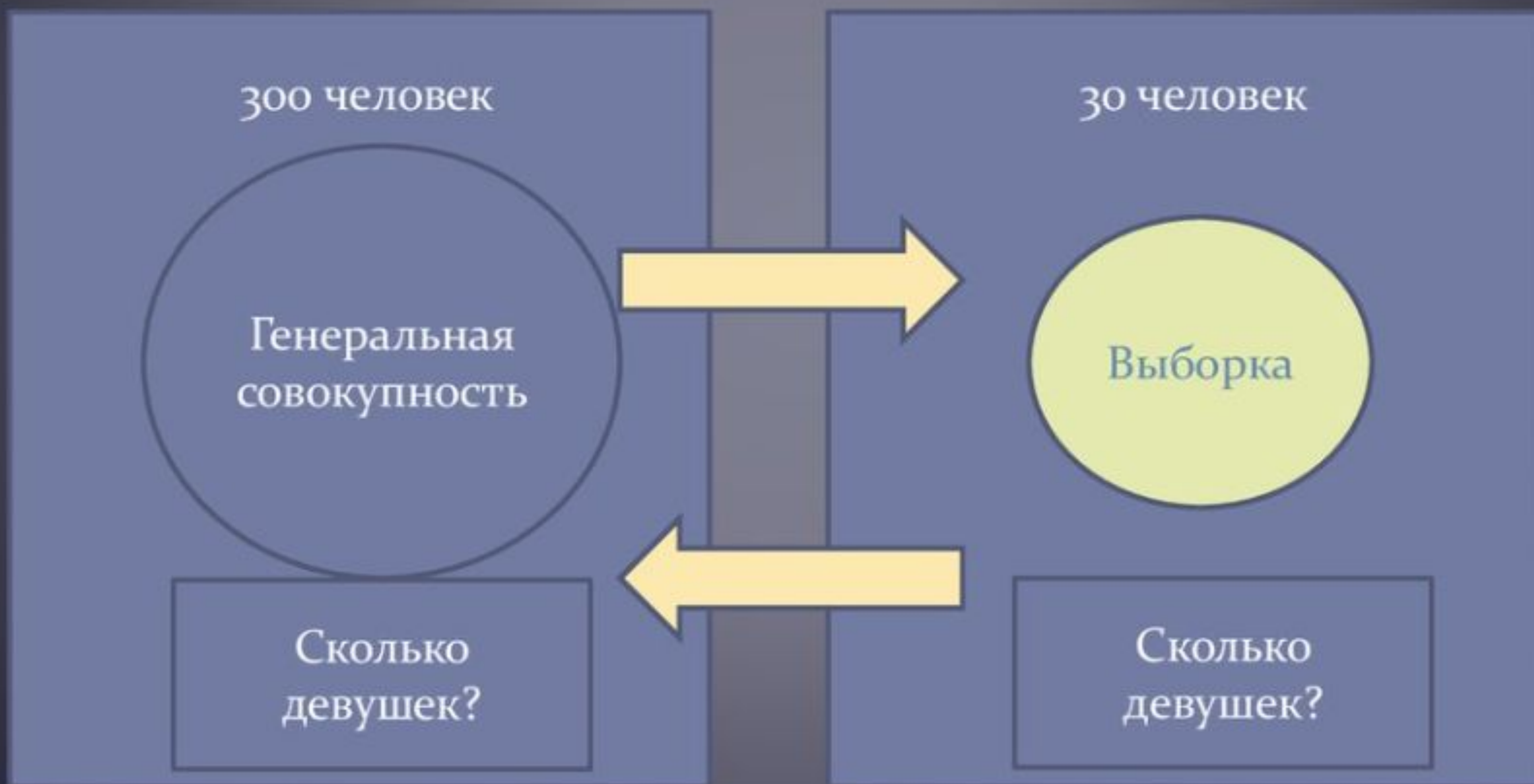
Основы математической статистики

- Предметом математической статистики является изучение случайных величин по результатам наблюдений.
- Задачи:
- 1. упорядочить данные
- 2. оценить характеристики наблюдаемой величины
- 3. проверить статистическую гипотезу
- Говорят, что «математическая статистика – это теория принятия решения в условиях неопределенности».

Генеральная совокупность и выборка



выборка

- Повторная
- Бесповторная

Способ отбора

- Простой
- Типический
- Механический
- Серийный

Пусть из генеральной совокупности извлечена выборка, причем x_1 наблюдалось n_1 раз, x_2 – n_2 раз, x_k – n_k раз и $\sum n_i = n$ – объем выборки. Наблюдаемые значения x_i называют ***вариантами***, а последовательность вариант, записанных в возрастающем порядке, – ***вариационным рядом***. Числа наблюдений n_i называют ***абсолютными частотами***, а их отношения к объему выборки $n_i / n = w_i$ – ***относительными частотами*** или ***частностями***.

Соответствие, установленное между наблюдаемыми вариантами и их частотами (абсолютными или относительными), *называют* ***статистическим распределением.***

При этом должны выполняться два ***условия нормировки***:

1) $n_1 + n_2 + \ldots + n_k = n$ (объем выборки);

2) $w_1 + w_2 + \ldots + w_k = 1$.

Удобной формой записи статистического распределения является таблица. В верхней строке таблицы записывают последовательность вариант, в нижней – соответствующие им частоты (абсолютные или относительные).

Пример 1. Имеются данные о количестве дежурств сотрудниками кафедры за месяц. Произведена выборка объемом $n = 15$:

3 0 5 7 4 3 1 9 5 3 4 4 2 8 5.

Составить статистический вариационный ряд распределения частот (абсолютных и относительных).

Решение

1. Расположить значения выборки в возрастающем порядке:

0 1 2 3 3 3 4 4 4 5 5 5 7 8 9.

Имеем девять различных значений.

2. Найти абсолютные частоты появления каждого значения выборки:

$n_1 = 1$, $n_2 = 1$, $n_3 = 1$, $n_4 = 3$, $n_5 = 3$, $n_6 = 3$, $n_z = 1$, $n_8 = 1$, $n_9 = 1$.

Проверить первое условие нормировки:

$$n = \sum_{i=1}^{9} n_i = 1+1+1+3+3+3+1+1+1=15$$

3. Вычислить относительные частоты появления каждого значения выборки по формуле $w_i = n_i \, / \, n$:

$w_1^* = 1/15$, $w_2 = 1/15$, $w_3 = 1/15$, $w_4 = 3/15$, $w_5 = 3/15$, $w_6 = 3/15$, $w_7 = 1/15$, $w_8 = 1/15$, $w_9 = 1/15$.

Проверить второе условие нормировки:

.

$$W = \sum_{i=1}^{9} \mathbf{w}_i^* = \frac{1}{15}+\frac{1}{15}+\frac{1}{15}+\frac{3}{15}+\frac{3}{15}+\frac{3}{15}+\frac{1}{15}+\frac{1}{15}+\frac{1}{15}=1$$

4. Внести полученные данные в таблицу:

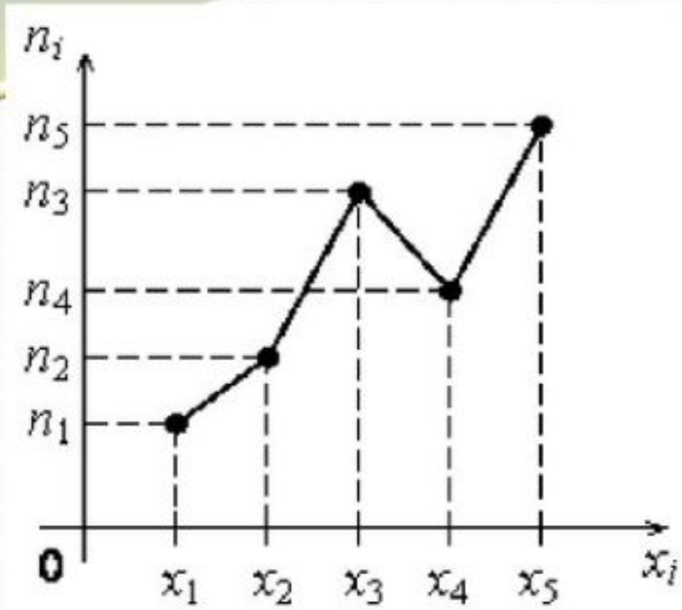
X_i	0	1	2	3	4	5	7	8	9
n_i	1	1	1	3	3	3	1	1	1
w_i	1/15	1/15	1/15	3/15	3/15	3/15	1/15	1/15	1/15

ПОЛИГОН

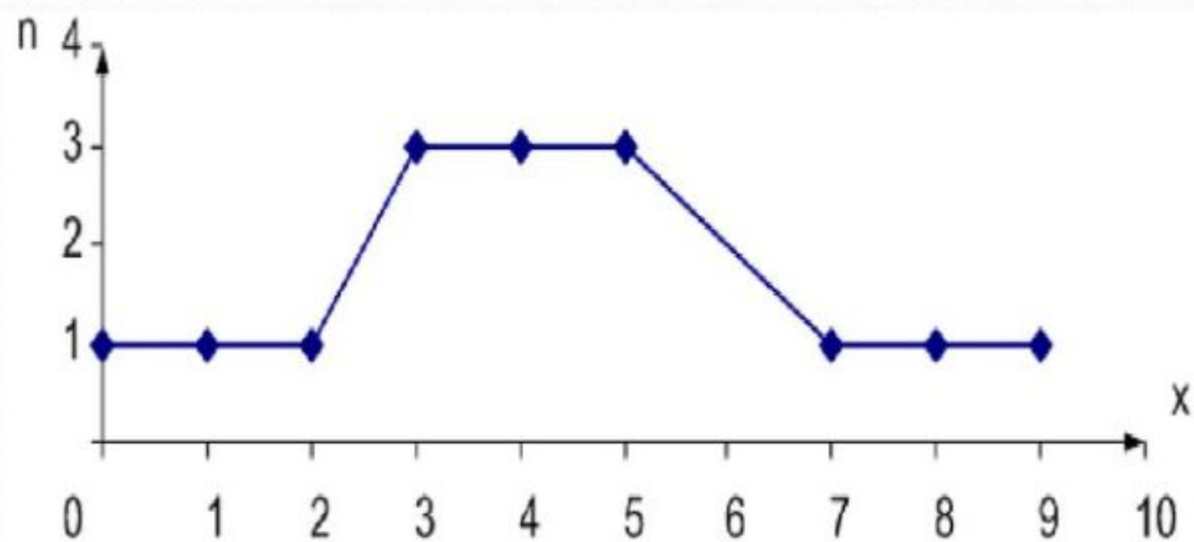
Для геометрического изображения такого статистического распределения служит *полигон частот* или *полигон относительных частот*.

Полигоном частот называют *ломаную линию, отрезки которой соединяют точки* $(x_1;n_1)$, $(x_2;n_2)$, …, $(x_k;n_k)$. Для построения полигона частот на оси абсцисс откладывают варианты x_i, а на оси ординат – соответствующие им частоты n_i.

Полигоном относительных частот называют *ломаную линию, отрезки которой соединяют точки* $(x_1; w_1)$, $(x_2; w_2)$, … , $(x_k; w_k)$. Для построения полигона частот на оси абсцисс откладывают варианты x_i, а на оси ординат – соответствующие им относительные частоты p_i



n_i
n_5
n_3
n_4
n_2
n_1
0
x_1
x_2
x_3
x_4
x_5
x_i



n
4
3
2
1
x
0
1
2
3
4
5
6
7
8
9
10

Гистограмма

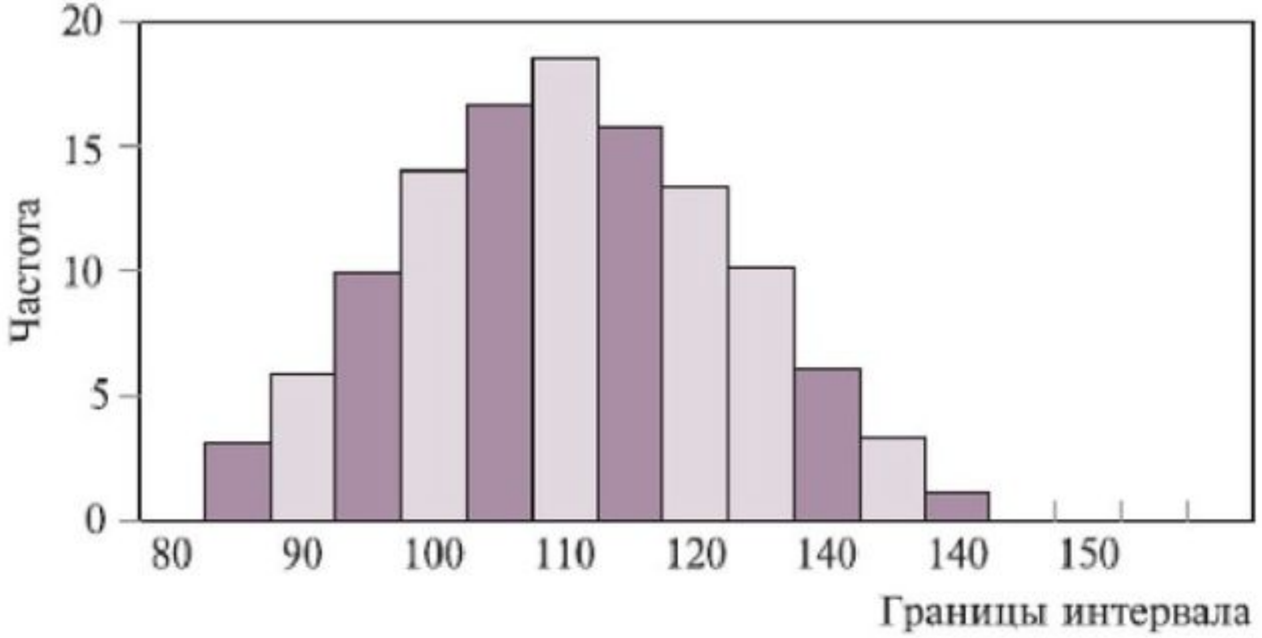
Гистограммой частот называется ступенчатая фигура, основанием i-го прямоугольника которой являются частичные интервалы длиною Δ_i, и высотой n_i.

Для построения гистограммы частот на оси абсцисс откладывают частичные интервалы, а над ними проводят отрезки, параллельные оси абсцисс на расстоянии n_i .

В практике для удобства вычислений обычно используют ряды с равными интервалами (Δ), которые называют шагом интервала.

Гистограммой относительных частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиною Δ_i, а высоты равны отношению w_i.

Гистограмма нормального распределения



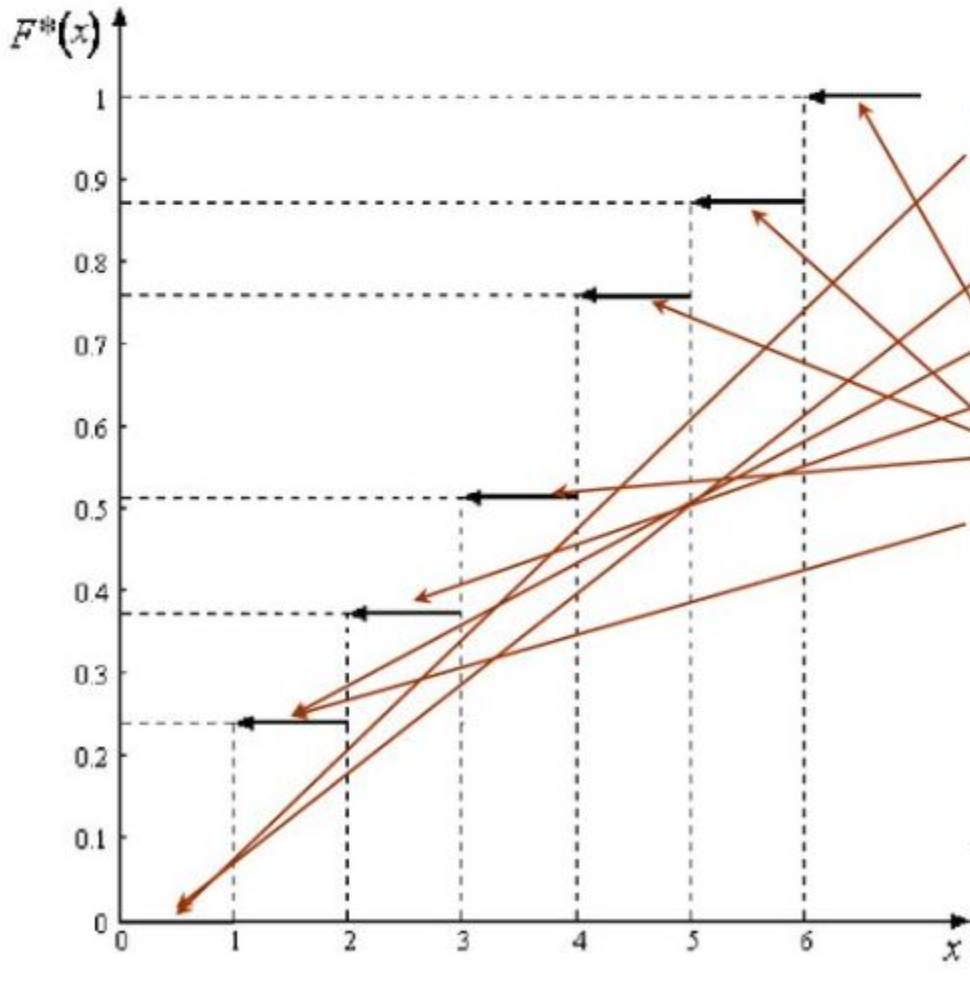
- **Эмпирической функцией распределения** (функцией распределения выборки) $F^*(x)$ называется **относительная частота** события $\{X < x\}$

$$F^*(x) = \frac{n_x}{n}$$

где n_x – число вариант, меньших x , n – объем выборки.

Построим

эмпирическую функцию распределения



$$F^*(x)=\begin{cases}0, & x\le 1,\\ 0{,}22, & 1<x\le 2,\\ 0{,}38, & 2<x\le 3,\\ 0{,}51, & 3<x\le 4,\\ 0{,}75, & 4<x\le 5,\\ 0{,}87, & 5<x\le 6,\\ 1, & x>6.\end{cases}$$

числовые характеристики выборки

- **Выборочное среднее** – *среднее арифметическое значений выборки*

$$\overline{x}_B = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{k} n_i \cdot x_i = \sum_{i=1}^{k} w_i \cdot x_i$$

числовые характеристики выборки

- **Выборочная мода** Mo^* – наиболее вероятное значение в выборке (*варианта с наибольшей частотой*).
- **Выборочная медиана** Me^* – значение случайной величины, *приходящееся на середину вариационного ряда*. Если объем выборки

  четен, *n=2m*, то, $Me^* = \dfrac{x_m + x_{m+1}}{2}$

  если нечетен, *n=2m+1*, то $Me^* = x_{m+1}$

числовые характеристики выборки

- ***Выборочная дисперсия*** *– среднее значение квадрата отклонения* $x_i - \overline{x}_B$

$$D_B = \frac{1}{n}\sum_{i=1}^{k} n_i \cdot \left(x_i - \overline{x}_B\right)^2 = \sum_{i=1}^{k} w_i \cdot \left(x_i - \overline{x}_B\right)^2$$

Эта формула может быть преобразована к виду

$$D_B = \frac{1}{n}\sum_{i=1}^{k} n_i \cdot x_i^{\,2} - \left(\overline{x}_B\right)^2 = \sum_{i=1}^{k} w_i \cdot x_i^{\,2} - \left(\overline{x}_B\right)^2 = \overline{x^2} - \left(\overline{x}\right)^2$$

$$\boxed{\overline{x^2} = M^*\left(x^2\right) \qquad \overline{x} = M^*\left(x\right)}$$

числовые характеристики выборки

- ***Выборочное среднее квадратическое отклонение*** $\sigma_B = \sqrt{D_B}$
- ***Исправленная выборочная дисперсия*** $$s^2 = \frac{n}{n-1} D_B = \frac{1}{n-1} \sum_{i=1}^{k} n_i \cdot \left(x_i - \overline{x}_B \right)^2$$
- ***Исправленное выборочное среднее квадратическое отклонение*** $s = \sqrt{s^2}$