Генеральной совокупностью будем называть совокупность всех мыслимых наблюдений результатов над случайной величиной, которые в принципе могли бы быть проведены в заданных условиях опыта (обозначают N).

Выборкой объема *n* называется конечный набор значений случайной величины, взятый в заданных условиях опыта.

Статистическое распределение выборки.

Пусть из генеральной совокупности N взята выборка объема n для изучения признака X, который принял значения x_1 в n_1 случаях, $x_2 - n_2$ раз, $..x_m - n_m$ раз. Значения $x_1, x_2, ..., x_m$ называется **вариантами признака** X **.**

Варианты x_i, записанные в возрастающем порядке с указанием частоты n_i, с которой они встретились в данной совокупности, образуют статистическое распределение выборки или **вариационный ряд.**

Количество наблюденных вариант $n_1, n_2, ..., n_m$ называется **рядом частот,** для которых выполняется следующее $\sum_{k=1}^{m} n_k = n$. Отношение частоты n_k, варианты x_k к объему выборки n называется **относительной частотой варианты (о**бозначается ω_k). Сумма всех относительных частот равна единице: $\sum_{k=1}^{m} \omega_k = 1$.

x_i	x_1	x_2	x_3	...	x_m
n_i	n_1	n_2	n_3	...	n_m
ω_i	ω_1	ω_2	ω_3	...	ω_m

$x_1 < x_2 < x_3 < ... < x_m$

Значение вариант и частот или относительных частот можно рассмотреть как координаты точек $M_1(x_1, n_1), M_2(x_2, n_2), ..., M_m(x_m, n_m)$ (1) или $M_1(x_1, \omega_1), M_2(x_2, \omega_2), ..., M_m(x_m, \omega_m)$ (2).

Полигоном частот называют ломаную, отрезки которой соединяют точки (1).

Полигоном относительных частот (частностей) называют ломаную, отрезки которой соединяю точки (2).

Полигоны частот и относительных частот являются аналогами плотности вероятности.

В некоторых случаях особенно для непрерывно распределенного признака X лучшим представлением случайной величины являются гистограммы частот или относительных частот. При непрерывном распределении признака весь интервал, в котором заключаются все наблюденные значения $[x_{\min}, x_{\max}]$ ***называемый размахом выборки*** разбивают на ряд частичных интервалов длинны h и находят n_i – сумму частот вариант попавших в i - ый интервал.

Гистограммой частот называют ступенчатую фигуру состоящую из прямоугольников основаниями которых служат частичные интервалы длины h, а высоты – отношению ${n_i}/{h}$ (плотность частоты), $h = x_i - x_{i-1}$.

Площадь частичного i-ого прямоугольника равняется $h\dfrac{n_i}{h} = n_i$ – сумме частот вариант попавших в i - вариант.

Площадь гистограммы частот равна сумме всех частот, т.е. объему выборки n.

Гистограммой относительных частот называют ступенчатую фигуру состоящую из прямоугольников с основанием h и высотами ${\omega_i}/{h}$ (плотность относительной частоты).

Площадь частичного i-ого прямоугольника равна $h\dfrac{\omega_i}{h} = \omega_i$.

Площадь гистограммы относительных частот равна сумме всех отсительных частот, т.е. 1.

Для построения гистограммы частот промежуток вариант $[x_{\min}, x_{\max}]$ называемый размахом выборки разбивают на несколько отрезков равной длинны h, затем подсчитывают сумму частот значений вариант признака X принадлежит каждому из полученных отрезков.

<u>Эмпирическая функция распределения и её свойства.</u>

Эмпирической функцией распределения (или функцией распределения выборки) называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X < x$. По определению, $F^*(x) = {n_x}/{n}$, где n_x – число вариант, меньших x; n – объем выборки.

Свойства эмпирической функции:

1. Значение эмпирической функции принадлежит отрезку $0 \le F^*(x) \le 1$.
2. $F^*(x)$ - неубывающая функция.
3. $F^*(x) = 0, x \le x_1$

$F^*(x) = 1, x > x_m$, где x_1 - наименьшая варианта, x_m - наибольшая варианта.

Тема Числовые характеристики статистического распределения выборки.

1. **Выборочным средним** называется среднее арифметическое вариант выборки с учетом их частот:

$$x_B = \frac{1}{n}\sum_{i=1}^{m} n_i x_i$$

$$x_B = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2. **Модой** называется значение варианты, имеющее наибольшую частоту. Для дискретного ряда мода определяется по частотам вариант и соответствует варианте с наибольшей частотой.
3. **Медиана**, это такое значение варьирующего признака, которое приходится на середину упорядоченного вариационного ряда. Для дискретного ряда:

$$M_e = \begin{cases} x_{\frac{n+1}{2}}, & n - нечетное \\ \frac{1}{2}\left[x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right], & n \text{ - } четное \end{cases}.$$

4. **Выборочной дисперсией** D_B называют среднее квадратов отклонения вариант от выборочного среднего с учётом соответствующих относительных частот:

$$D_B = \frac{1}{n}\sum_{i=1}^{m} n_i \left(x_i - \overline{x_B} \right)^2 = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - \left(\overline{x_B} \right)^2$$

5. Выборочным среднеквадратическим отклонением (стандартом) называется $\sqrt{D_B}$

$$\sigma_B = \sqrt{D_B} \ .$$

6. **Коэффициентом вариации** V называется отношение выборочного среднего квадратического отклонения σ_B к выборочному среднему x_B, т.е.

$$V = \frac{\sigma_B}{x_B} \cdot 100\%$$