Содержание

1 Математическая статистика

1.1 Вариационные ряды и их характеристики

1.1.1 Генеральная и выборочная совокупности, объём выборки.

Рассмотрим постановку задачи математической статистики: по результатам наблюдения за некоторой случайной величиной ξ требуется сделать выводы о неизвстном законе распределения этой величины $\mathcal{L}(x, \Theta)$ либо о неизвестных парамерах $\Theta_1, \ldots, \Theta_n$ известного распределения.

Пусть ξ — случайная величина с некоторой (теоретической) функцией распределения $F_\xi(x) = P\{\xi < x\}, \quad x \in R$.

Определение:

Совокупность n независимых одинаково распределённых случайных величин $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ называется выборкой (выборочной совокупностью), извлечённой из распределения случайной величины ξ.

Определение:

Под генеральной совокупностью понимается множество всех возможных значений случайной величины ξ.

Определение:

Объёмом совокупности называется количество всех её элементов, объё выборки или выборочной совокупности обозачается n, генеральной совокупности — N.

1.1.2 Варианционный ряд, варианта, частота. Виды вариационных рядов. Гистограмма, полигон.

Определение:

Пусть $x_1, x_2, \ldots, x_n$ — выборка из генеральной совокупности значений.

Вариационным рядом называется последовательность $x_1^*, x_2^*, \ldots, x_n^*$ элементов выборки расположенных в порядке неубывания, т.е. $x_1^* \leq x_2^* \leq \cdots \leq x_n^*$.

x_i^* — варианта.

n_i — частота появления варинты x_i^* в выборке.

Определение:

Точечным вариационным рядом называется:

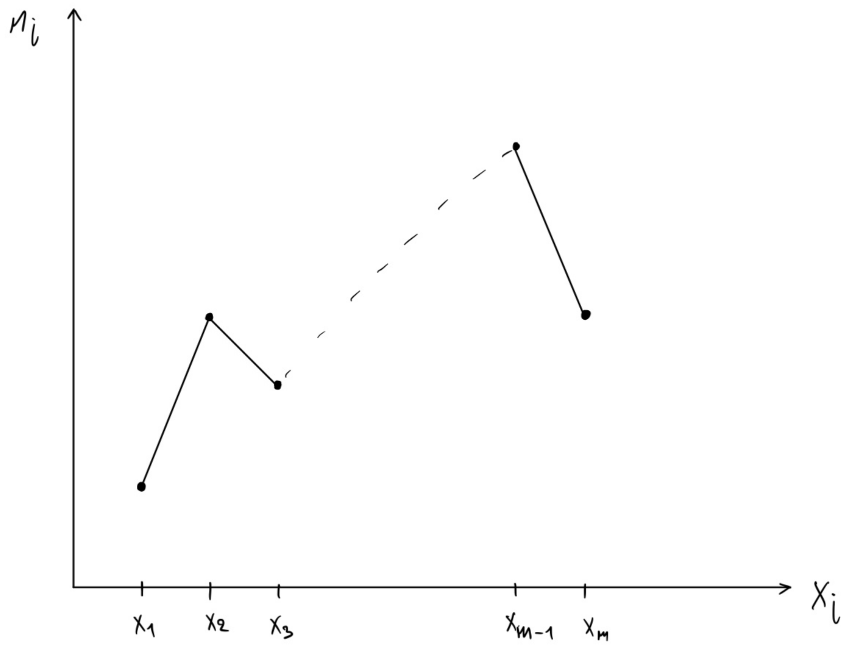
x_i	x_1	x_2	...	x_m
n_i	n_1	n_2	...	n_m

x_i — варианта, n_i — частота соответствующей варианты.

m — количество групп (различных вариант (вариант в таблице)).

$n = \sum_{i=1}^{m} n_i$, где n_i — объём выборки.

Для графического представления точечных вариационных рядов используется полигон частот — ломанная с вершинами в точках (x_i, n_i).



Определение:

Интервальным вариационным рядом называется:

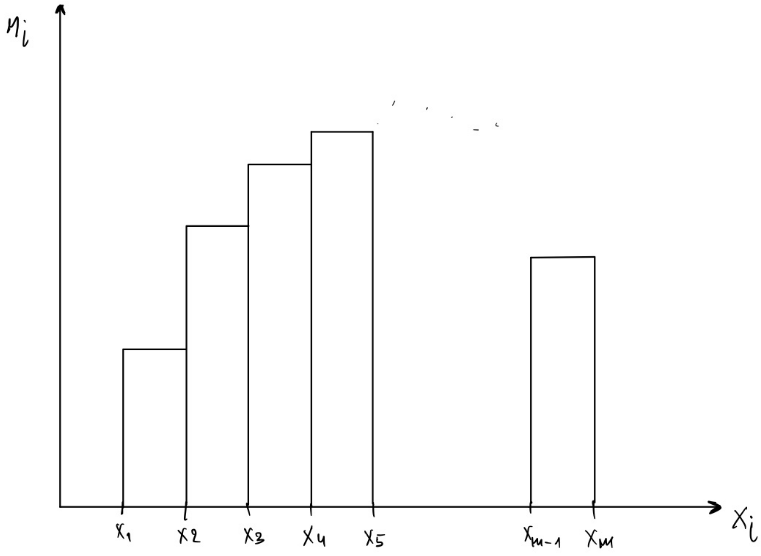
x_i	$[x_1, x_2]$	$(x_2, x_3]$	...	$(x_m, x_{m+1}]$
n_i	n_1	n_2	...	n_m

x_i — варианты, n_i — частота.

m — количество групп (интервалов).

$n = \sum_{i=1}^{m} n_i$, где n_i — объём выборки.

Для графического представления интервальных вариационных рядов используется гистограмма частот — фигура, составленная из прямоугольников, одной стороной которых служат интервалы $(x_i, x_{i+1}]$, а длина второй равна n_i.



1.1.3 Формулы числовых характеристик. Эмперическая функция распределения (ЭФР). Свойства ЭФР.

Определение:

Выборочным средним называется величина:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X_i}.$$

Если данные представлены в виде точечного или интервального вариационного ряда, то для вычисления используют формулу:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i * n_i,$$

где m — количество групп в точечном или интервалов в интервальном вариационном ряду, n_i — частота, т.е. количество элементов выборки, принадлежащихх i-той группе или i-тому интервалу, x_i — варианта для точечного ряда и середина i-того интервала для интервального ряда.

Определение:

Выборочной дисперсией (смещённой) называется величина:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \overline{x})^2.$$

Она характеризует среднее из квадратов отклонений наблюдаемой величины от выборочного среднего. Величина $S = \sqrt{S^2}$ называется выборочным средним квадратическим отклонением (смещённым) величин выборки от выборочного среднего.

Если данные представлены в виде точечного или интервального вариационного ряда, то для вычисления используют формулу:

$$S^2 = \frac{1}{n}\sum_{i=1}^{m}(x_i - \overline{x})^2 n_i.$$

или

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

Здесь m — количество групп в точечном или интервалов в интервальном вариационных рядах, n_i — частота. т.е. количество элементов выборки, принадежащих i-той группе или i-тому интервалу, x_i — варианта для точечного ряда и середина i-того интервала для интервального ряда.

Определение:

Выборочной дисперсией (несмещённой) называется величина:

$$\overline{\sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i\overline{x})^2.$$

Аналогично, величина $\overline{\sigma} = \sqrt{\overline{\sigma}^2}$ называется выборочным несмещённым средним квадратическим отклонением.

Очевидно, что смещённая и несмещённая выборочные дисперсии связаны формулой:

$$\overline{\sigma}^2 = \frac{n}{n-1}S^2.$$

1.1.4 Эмперическая функция распределения ЭФР. Свойства ЭФР.

Эмперическая функция распределения ЭФР

Эмперической функцией распределения (ЭФР) называется функция

$$\tilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} e(x - \mathbf{X}_i),$$

где $e(x) = 1$, при $x > 0 \quad e(x) = 0$, при $x \leq 0$.

Таким образом, если $\mathbf{X}_i$, то $e(x) = 1$, если $\mathbf{X}_i \geq x$, то $e(x) = 0$, а сумма $e(x - \mathbf{X}_i)$ будет равна количеству элементов выборки, которые приняли значение, строго меньше некоторого $x \in R$.

Пусть $x_1, x_2, \ldots, x_n$ — реализация выборки $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, т.е. наблюдавшиеся значения случайной величины ξ.

Обозначим $\mu(x)$ — число элементов выборки, строго меньших $x \in R$. Тогда эмпирическая функция распределения $\tilde{F}_n(x)$ может быть определена как

$$\tilde{F}_n(x) = \frac{\mu(x)}{n}.$$

Свойства ЭФР:

1. $0 \leq \tilde{F}_n(x) \leq 1$, т.к. $0 \leq \mu(x) \leq n$;
2. неубывающая непрерывная слева функция;
3. $\tilde{F}_n(x)$ — ступенчатая функция для всех типов распределений;
4. $\tilde{F}_n(x)$ сходится по распределению к $F_\xi(x)$.