

Статистические методы обработки информации

Лекция 4

Что такое статистика?

- ▣ **Математическая статистика** – это раздел математики, занимающийся установлением закономерностей, которым подчинены массовые случайные явления, на основе обработки статистических данных, полученных в результате наблюдений.



Основные задачи математической статистики

- определение способов сбора и группировки этих статистических данных;
- разработка методов анализа полученных данных в зависимости от целей исследования.



Основные понятия математической статистики

- **Случайная величина** – величина, принимающая то или иное числовое значение в испытании (заранее неизвестно, какое именно).
- **Дискретные**
 - Множество значений конечно или счетно
 - СВ-число очков на верхней грани кубика
- **Непрерывные**
 - Значения СВ принадлежат некоторому промежутку, конечному или бесконечному
 - Рост студентов группы



Основные понятия математической статистики

Предположение. Выбран и измеряется в последовательности экспериментов некоторый числовой признак (случайная величина) X .

Определение. *Генеральная совокупность* – все возможные значения признака в результате всех возможных мыслимых экспериментов.

Как правило, генеральная совокупность нам недоступна, а если и может быть указана, то только теоретически.

- ▣ *Выборка* – часть генеральной совокупности, отбираемая специальным образом и исследуемая с целью получения выводов о свойствах генеральной совокупности.
 - ▣ *Объем выборки* – количество значений в выборке.
-



Выборка

Повторная

- каждый отобранный объект перед выбором следующего возвращается в генеральную совокупность

Бесповторная

- отобранный объект в генеральную совокупность не возвращается



Репрезентативность

- - это главный критерий надежности выборки.

Выборка репрезентативна, если

- каждый объект генеральной совокупности имеет одинаковую вероятность быть отобранным;
- объекты выбирают независимо друг от друга.
- объекты выбираются случайно.



Вариационный ряд и его характеристики

Определение. *Вариационным рядом* называются значения признака выборки, расположенные в порядке возрастания:

$$x_1 \leq x_2 \leq \dots \leq x_n.$$

Определение. *Медианой* M_e вариационного ряда называется значение, расположенное в его середине (если n – нечетное, то в точности в середине, если n – четное, то либо два соседних в середине, либо их полусумма).

Определение. *Модой* M_o вариационного ряда называется значение, которое встречается в нем чаще всего. Если таких значений несколько, то ряд называется *полимодальным*.

Определение. *Вариационным размахом* R вариационного ряда называется разность между наибольшим и наименьшим значениями признака в ряде.

Пример 1.

- При проведении 20 серий из 10 бросков игральной кости число выпадений шести очков оказалось равным 1, 1, 4, 0, 1, 2, 1, 2, 2, 0, 5, 3, 3, 1, 0, 2, 2, 3, 4, 1.
- Составьте вариационный ряд. Определите моду и медиану вариационного ряда.



Статистический ряд

Пусть дана выборка. Рассмотрим только *различные* значения признака в выборке и обозначим их в порядке возрастания через x_1, \dots, x_k . Они называются **вариантами**. Варианта может встречаться несколько раз. Количество появлений значения x_i в вариационном ряде называется *частотой* этого значения и обозначается n_i . **Сумма всех частот равна объему выборки:**

$$n_1 + \dots + n_k = n.$$

Определение. *Статистическим рядом* называется ранжированный (то есть упорядоченный по возрастанию) набор различных значений признака в выборке (вариант), вместе с их весами (частотами).

x_i	x_1	\dots	x_k
n_i	n_1	\dots	n_k

Таблица относительных частот

Определение. Отношение частот значений статистического ряда к объему выборки называются *частостями* или *относительными частотами* и обозначаются через w_i :

$$w_i = \frac{n_i}{n}, \quad i = 1, 2, \dots, k.$$

Статистический ряд удобно представлять и в виде таблицы частостей:

x_i	x_1	\dots	x_k
w_i	w_1	\dots	w_k

Упражнения. 1) Чему равна сумма частостей? 2) Как по статистическому ряду сразу указать моду вариационного ряда?

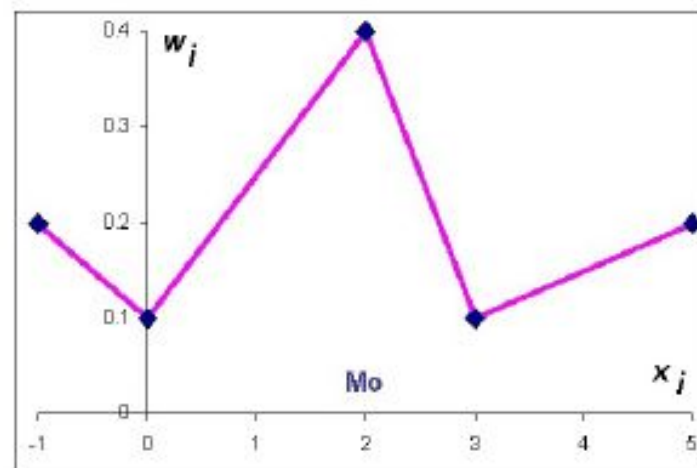


Графическое представление статистических данных

Определение. *Полигоном частот* статистического ряда называется ломаная, составленная по точкам (x_i, n_i) , то есть по оси абсцисс откладываются значения признака x_1, \dots, x_k , а по оси ординат – их частоты n_1, \dots, n_k . Если вместо частот по оси ординат откладываются частоты, то ломаная называется *полигоном частостей*.

Пример. Полигон частостей. Статистический ряд задан таблицей:

x_i	-1	0	2	3	5
w_i	0.2	0.1	0.4	0.1	0.2



Пример 1.

- При проведении 20 серий из 10 бросков игральной кости число выпадений шести очков оказалось равным 1, 1, 4, 0, 1, 2, 1, 2, 2, 0, 5, 3, 3, 1, 0, 2, 2, 3, 4, 1.
- Составьте статистический ряд, таблицу относительных частот.
- Постройте полигон относительных частот.



Группировка данных

- В случае *большого объема выборки* рассматривают значения признака в пределах определенного интервала.
- **Интервальное распределение частот** – это таблица, состоящая из некоторого количества интервалов равной длины h , на которые делится весь диапазон изменения признака, и соответствующих этим интервалам частот.
$$h = \log_2 n + 1$$
- **Интервальная частота m_h** - количество случаев, в которых признак принял значения, входящие в некоторый интервал.



Гистограмма

- *Графическая иллюстрация интервального распределения*
- ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной h , а высотами – отрезки длиной n_i / h (гистограмма частот) или w_i / h (гистограмма относительных частот).
- В первом случае площадь гистограммы равна объему выборки, во втором – единице.



Пример 2.

- Имеются данные о продолжительности 20 разговоров в минутах. Получите интервальное распределение частот. Постройте гистограмму частот.

11	29	6	33	14
21	18	17	22	38
31	22	27	19	22
23	26	39	34	27



Характеристики вариационного ряда

Определение. Пусть дана выборка объема n . Если $x_1 \leq \dots \leq x_n$ – вариационный ряд, то величина

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

называется *средней арифметической* (вариационного ряда) или *выборочным средним*.

Следствие. Для статистического ряда $(x_1, n_1), \dots, (x_k, n_k)$ выборочное среднее вычисляется по формуле

$$\bar{x} = \frac{\sum_{i=1}^k n_i x_i}{\sum_{i=1}^k n_i} = \sum_{i=1}^k w_i x_i.$$



Характеристики вариационного ряда

Определение. Выборочная дисперсия статистического ряда определяется по формуле

$$s^2 = \frac{\sum_{i=1}^k n_i (x_i - \bar{x})^2}{n} = \sum_{i=1}^k w_i (x_i - \bar{x})^2 = \left(\sum_{i=1}^k w_i x_i^2 \right) - \bar{x}^2.$$

Определение. Среднее квадратическое отклонение статистического ряда определяется как корень квадратный из выборочной дисперсии:

$$s = \sqrt{s^2}.$$



Спасибо за внимание!