

Лекция 13

Математическая статистика (МС)

§ 1. Основные понятия выборочного метода

Статистика – наука о придании смысла данным. МС исследует закономерности массовых случайных явлений на основе результатов наблюдений или испытаний.

Первая задача МС – указать способы сбора и группировки статистических данных, полученных в результате наблюдений или экспериментов.

Вторая задача –разработать методы анализа статистических данных в зависимости от целей исследования. К ним относятся:

- 1) Оценка неизвестной вероятности события, оценка неизвестной функции распределения (закона распределения), оценка параметров распределения, вид которого известен, оценка зависимости СВ от одной или нескольких СВ.
- 2) Проверка статистических гипотез о виде неизвестного распределения или о величине параметров распределения, вид которого известен.

Таким образом, задача МС – создание методов сбора и обработки статистических данных для получения научных и практических выводов.

Пусть требуется изучить совокупность однородных объектов относительно некоторого количественного или качественного признака, характеризующего эти объекты.

Генеральной совокупностью называется множество всех однородных объектов, подлежащих изучению.

Выборочной совокупностью (или выборкой) называется часть генеральной совокупности, отобранная для изучения.

Цель МС – оценить характеристики генеральной совокупности по выборочным данным на основании методов теории вероятностей. Полученные выводы называются *статистическими*.

Объемом совокупности (выборочной или генеральной) называется число объектов в этой совокупности. Часто генеральная совокупность содержит конечное число объектов, но если это число достаточно велико, допустимо считать, что генеральная совокупность состоит из бесконечного множества объектов. Такое допущение оправдано тем, что увеличение объема генеральной совокупности практически не сказывается на результатах обработки данных выборки.

Существует несколько способов отбора данных в выборку:

повторной называют выборку, при которой отобранный объект перед отбором следующего возвращают в генеральную совокупность;

бесповторной называют выборку, при которой отобранный объект перед отбором следующего не возвращают в генеральную совокупность.

Часто используют бесповторный отбор.

Для того, чтобы по данным выборки можно было достаточно уверенно судить об интересующем нас признаке генеральной совокупности, необходимо. Чтобы объекты выборки правильно его представляли. Выборка должна правильно представлять пропорции генеральной совокупности, т.е. она должна быть **репрезентативной**.

В силу ЗБЧ можно утверждать, что выборка будет репрезентативной. Если ее осуществить случайно. Каждый объект выборки выбран случайно из генеральной совокупности, если все объекты имеют одинаковую вероятность попасть в выборку.

Если объем генеральной совокупности достаточно велик, то выборка составляет лишь незначительную часть этой совокупности и различие между повторной и бесповторной выборками стирается. В предельном случае, если генеральная совокупность бесконечна, это различие исчезает.

На практике применяют два способа отбора:

- 1) Отбор, не требующий расчленения генеральной совокупности. Сюда относят:
 - а) простой случайный бесповторный отбор,
 - б) простой случайный повторный отбор.
- 2) Отбор, при котором генеральная совокупность разбивается на части. Сюда относят:
 - а) типический отбор,
 - б) механический отбор,
 - в) серийный отбор.

Простым случайным отбором называют отбор, при котором объекты генеральной совокупности по одному извлекаются из генеральной совокупности.

Типическим называют отбор, при котором объекты отбирают не из всей генеральной совокупности, а из каждой типической ее части. Например, при исследовании доходов населения его можно предварительно классифицировать по социальному статусу (чиновники, служащие, рабочие и т.д.).

При *механическом* отборе элементы из генеральной совокупности отбираются в выборку по заранее установленному правилу.

Серийным называют отбор, при котором объекты из генеральной совокупности выбираются не по одному, а сериями.

На практике применяют комбинированный отбор, при котором сочетают описанные выше способы. Например, разбивают генеральную совокупность на серии одинакового объема, простым случайным отбором выбирают несколько серий, затем из каждой сери, простым случайным отбором выбирают объекты.

§ 2. Статистическое распределение выборки

Пусть из генеральной совокупности извлечена выборка, причем значение x_1 наблюдалось n_1 раз, x_2 - n_2 раз, ..., x_k - n_k раз, $\sum_i n_i = n$ - объем выборки.

Наблюдаемые значения x_i называют **вариантами**, последовательность вариантов в порядке возрастания – **вариационным рядом**.

Числа наблюдений n_i называют **частотами**, а их отношения к объему выборки $w_i = \frac{n_i}{n}$ - **относительными частотами** или **частостями**.

Статистическим распределением выборки называют перечень вариантов и соответствующих им частот и относительных частот. Оно также может иметь вид последовательности интервалов и соответствующих им частот (в качестве такой частоты принимают сумму частот вариантов, попавших в этот интервал).

Размах выборки $R = x_{\max} - x_{\min}$.

Графически статистический ряд можно представить в виде **полигона** (многоугольника) частот (относительных частот) – ломаной, соединяющей точки (x_i, n_i) или (w_i, n_i) . Полигон является аналогом многоугольника распределения дискретной СВ.

§ 3. Эмпирическая функция распределения

По статистическому ряду можно **построить эмпирическую функцию распределения**, значения которой вычисляются по формуле

$$F^*(x) = \sum_{x_i < x} w_i.$$

Суммирование ведется по всем индексам i , для которых $x_i < x$.

Эмпирическая функция распределения обладает всеми свойствами теоретической функции распределения.

§ 4. Полигон и гистограмма

Полигон частот (относительных частот) – ломаная, соединяющая точки (x_i, n_i) или (w_i, p_i) .

Если рассматривают интервальный статистический ряд, то строят гистограмму.

Гистограмма – это ступенчатая фигура, состоящая из прямоугольников. Длина основания каждого прямоугольника равна длине интервала $h_i = x_i - x_{i-1}$. Высота прямоугольников

$$n_i^* = \frac{n_i}{h_i} \text{ (гистограмма частот),}$$

или

$$p_i^* = \frac{w_i}{h_i} \text{ (гистограмма относительных частот).}$$