Полигон и гистограмма. Эмпирическая функция распределения

КАЛАБУХОВА Галина Валентиновна
К.социол.н., доцент

Вопросы темы

- Задачи математической статистики.
- Эмпирическая функция распределения
- Полигон и гистограмма.

Задачи математической статистики

- ***Первая задача математической статистики*** - указать способы сбора и группировки статистических сведений, полученных в результате наблюдений или в результате специально поставленных экспериментов

- ***Вторая задача математической статистики*** - разработать методы анализа статистических данных в зависимости от целей исследования. Сюда относятся:
  - оценка неизвестной вероятности события; оценка неизвестной функции распределения; оценка параметров распределения, вид которого известен; оценка зависимости случайной величины от одной или нескольких случайных величин и др.;
  - проверка статистических гипотез о виде неизвестного распределения или о величине параметров распределения, вид которого известен.

- Современная математическая статистика разрабатывает способы определения числа необходимых испытаний до начала исследования (планирование эксперимента), в ходе исследования (последовательный анализ) и решает многие другие задачи. Современную математическую статистику определяют как науку о принятии решений в условиях неопределенности.

Выборочный метод

Генеральная и выборочная совокупности

Задача

Пусть требуется изучить совокупность однородных объектов относительно некоторого качественного или количественного признака, характеризующего эти объекты. Иногда проводят сплошное обследование: обследуют каждый из объектов совокупности относительно признака, которым интересуются. На практике, однако, сплошное обследование применяют сравнительно редко. Например, если совокупность содержит очень большое число объектов, то провести сплошное обследование физически невозможно. Если обследование объекта связано с его уничтожением или требует больших материальных затрат, то проводить сплошное обследование практически не имеет смысла. В таких случаях случайно отбирают из всей совокупности ограниченное число объектов и подвергают их изучению

Определения

- ***Выборочной совокупностью*** или просто *выборкой* называют совокупность случайно отобранных объектов.
- ***Генеральной совокупностью*** называют совокупность объектов, из которых производится выборка.
- ***Объемом*** совокупности (выборочной или генеральной) называют число объектов этой совокупности.

Например, если из 1000 деталей отобрано для обследования 100 деталей, то объем генеральной совокупности N=1000, а объем выборки n = 100.

Эмпирическая функция распределения

Определения

Пусть из генеральной совокупности извлечена выборка, причем x_1 наблюдалось n_1 раз, x_2 - n_2 раз, x_k- n_k раз и - объем выборки.

Наблюдаемые значения x_i - называют ***вариантами***, а последовательность вариант, записанных в возрастающем порядке, - ***вариационным рядом***.

Числа наблюдений называют *частотами*, а их отношения к объему выборки $n_i /n = W_i$ - ***относительными частотами***.

Статистическим распределением выборки называют перечень вариант и соответствующих им частот или относительных частот

табличное задание выборки

x_i	x_1	x_2	…	x_n
w_i	w_1	w_2	…	w_n

Пусть известно статистическое распределение частот количественного признака X. Введем обозначения: n_x - число наблюдений, при которых наблюдалось значение признака, меньшее x; n - общее число наблюдений (объем выборки). Ясно, что относительная частота события $X < x$ равна n_x/n. Если x изменяется, то, вообще говоря, изменяется и относительная частота, т.е. относительная частота n_x/n есть функция от x.

Определение

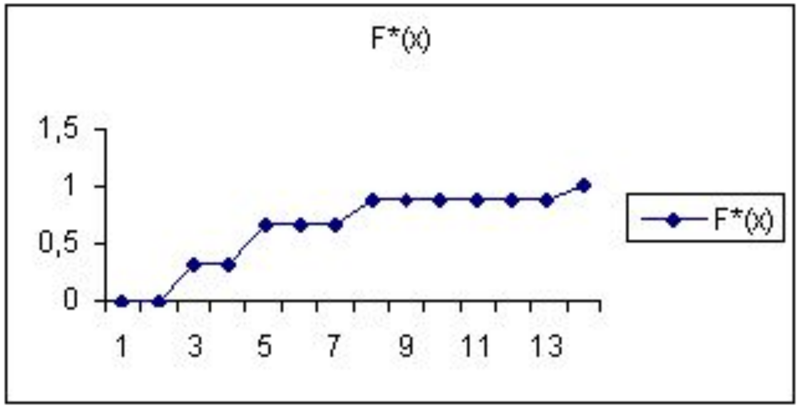
Эмпирической функцией распределения (функцией распределения выборки) называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X < x$.

Итак, по определению, $F^*(x) = n_x/n$,

где n_x - число вариант, меньших x;

n - объем выборки

эмпирическая функция распределения



Свойства эмпирической функции распределения

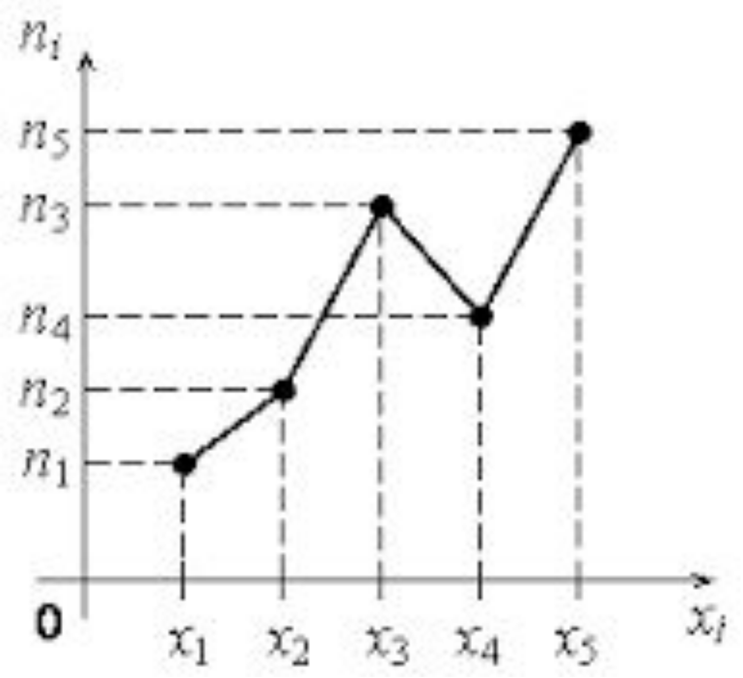
1. значения эмпирической функции принадлежат отрезку [0, 1];
2. $F^*(x)$ - неубывающая функция;
3. если x_1 - наименьшая варианта, то $F^*(x) = 0$ при $x \leq x_1$; если x_k - наибольшая варианта, то $F^*(x)=1$ при $x>x_k$.

Полигон и гистограмма

Определение

Полигоном частот называют ломаную, 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. Точки $(x_i; n_i)$ соединяют отрезками прямых и получают полигон частот.



Определение

Полигоном относительных частот *называют ломаную, отрезки которой соединяют точки* $(x_1; W_1), (x_2; W_2), \ldots, (x_k; W_k)$

Для построения полигона относительных частот на оси абсцисс откладывают варианты x_i а на оси ординат- соответствующие им относительные частоты W_i. Точки $(x_i; W_i)$ соединяют отрезками прямых и получают полигон относительных частот.

Определение

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиною *h,* а высоты равны отношению n_i/h (плотность частоты)