Лекция 11. Элементы математической статистики. Выборочный метод.

§1. Задача математической статистики.

Установление закономерностей, которым подчинены массовые случайные явления, основано на изучении статистических данных – результатах наблюдений. Основная задача математической статистики – указать способы сбора и группировки статистических сведений. Другая задача математической статистики - разработка методов анализа статистических данных в зависимости от целей исследования.

§2. Понятие о выборочном методе исследования.

Существует метод сплошного обследования исследуемой совокупности объектов, т.е. каждый объект обследуется относительно признака, который нужен. На практике этот метод применяется редко. Этот метод неприменим в силу обстоятельств: если обследующих объектов большое количество, например хлопка в партии, или обследование материала связано с его уничтожением, например, ткань исследуется на прочность, в таких случаях проводить сплошное обследование не имеет смысла. Наиболее распространенный и применяемый часто на практике – выборочный метод.

Выборочный метод состоит в определении сводных характеристик (показателей) какой-либо статистической совокупности путем наблюдения не всех, а лишь части составляющих ее членов, взятых на выборку. Например, для определения среднего срока службы большой партии электрических лампочек отбирается сравнительно небольшая часть их и испытывается. Средний срок службы испытанных лампочек принимается за приближенное значение среднего срока службы лампочек во всей партии. Выбор n единиц из совокупности N объем должен быть «репрезентативным», т.е. должен производиться так, чтобы свойства членов, попавших в «выборку», правильно отражали соответствующие свойства всей совокупности. По закону больших чисел достаточно обширная выборка будет репрезентативной, если ее производить случайно: каждый объект выборки отобрал случайно из генеральной совокупности, если все объекты имеют одинаковую вероятность попасть в выборку. Так, из N объектов выбрать любые n можно с помощью жребия.

На практике применяются различные способы отбора, например, простой, случайный бесповторный отбор, простой случайный повтор, типические, механические, серийные и др.

Типические выборки заключаются в разбивке исследуемой совокупности на типические относительно некоторого признака группы и в отборе членов по определенным правилам из каждой такой группы. При этом необходимо, чтобы внутри каждой группы отбор проводили случайно.

Механические выборки заключаются в том, что исследуемая совокупность разбивается на группы механическим способом (например, каждый 10 или 20-й член совокупности).

Серийные выборки заключаются в том, что отбираются целые группы (серии) членов исследуемой сосвкупности и каждая группа подвергается сплошному обследованию (т.е. обследованию всех членов совокупности). Серийной выборкой пользуются тогда, когда обследуемый признак колеблется в различных сериях незначительно.

§3. Основная задача математической теории выборочного метода.

Определение 1. Все множество объектов, подлежащих контролю и исследованию, называются генеральной совокупностью.

Определение 2. Множество, случайным оборазом отобранных объектов, называется выборочной совокупностью.

Определение 3. Число объектов выборочной совокупности (или генеральной совокупности) называют объемом выборки (или генеральной совокупностью).

Например, если из 10000 деталей отобрано для контроля 100, то говорят N=10000 – генеральная совокупность, n=100 + объем выборки.
Обычно генеральная совокупность содержит конечное множество объектов. Но оно достаточно велико, поэтому при теоретических выводах объем генеральной совокупности часто предполагается бесконечным. Это оправдывается тем, что увеличение обема генеральной совокупности уже не сказывается на результатах отработки данных выборки.
Для объектов генеральной совокупности определяется некоторая числовая характеристика – это случайная величина ξ, принимающая на каждом объекте некоторое числовое значение. Сделав выборку, мы получаем ряд значений этой случайной величины $\xi - X_1, X_2, ..., X_n$. По этой последовательности значений нам следует приблизительно представить функцию распределения случайно величины ξ, ее математическое ожидание и дисперсию.
В 1933г. советский математик В.И. Гливенко была доказан основная теорема математической статистики, из которой следует правило для приближенного получения функции распределения случайной величины ξ.
Смысл ее в следующем: для любого действительного числа х обозначим $n(x)$ частот чисел x_n из выборки $x_1 ... x_n$, удовлетворяющих неравенству $x_n < x$.
Этим на всей числовой прямой определена функция $n(x)$. Положим $F^*(x) = \frac{n(x)}{n}$.
Эта функция называется функцией распределения выборки случайной величины ξ. Она идает приближенную функцию распределения $F(x)$ случайно величины ξ.
Пример. В результате выборки имеем
-3,+2,-1,-3,+5,-3,+2. Потсроить график эмпирической функции распределения.
Решение. $n = 7, x_1 = x_4 = x_6 = 3; \quad x_2 = x_7 = 2$.

$$n(x)=\begin{cases}0, & при \quad x<-3\\ 3, & -3<x<-1\\ 4, & -1\le x<2\\ 6, & 2\le x<5\\ 7, & x\ge 5\end{cases} \Rightarrow F^*(x)=\frac{n(x)}{n}=\begin{cases}0, & x\le -3\\ \frac{3}{7}, & -3<x\le -1\\ \frac{4}{7}, & \\ \frac{6}{7}, & \\ 1 & \end{cases}$$

График эмпирической функции $F^*(x)$ имеет вид

На примере видны основные особенности эмпирической функции распределения. Как и теоретическая, она не убывает, и ее значения $\in[0,1]$. Но эта формула всегда ступенчатая. Эмпирическая функция не зависит от того, в каком порядке сделана выборка, т.е. от того, в каком порядке идут числа в последовательности: $x_1x_2...x_n$.

Кроме эмпирической функции распределения и ее графика бывает полезно изобразить аналог плотности вероятности. Это делают двумя способами. Для каждого x_n подсчитывают частоту n_k. Откладывая на координату плоскости эти значения, проводят ломанную, которая называется полигон частот. Это график дает понятие о том, насколько часто встречается каждое значение. Вместо частоты $\frac{n_k}{n}$ и строится соответствующий полигон. Он отличен от предыдущего только изменением масштаба по оси ОУ.

Более наглядной представление о случайно величине дает гистограмма частот (относительных частот). Для ее построения весь промежуток $[X_{\min}, X_{\max}]$ разбивается на равные промежутки длины h. Для каждого из них подсчитывается число наблюденных значений, в него попавших. Если на промежутке Δl_S число значений n_S, то строится прямоугольник с основанием Δl_S и высотой $\frac{n_S}{n}$. Получается чертеж, называющийся гистограммой частот. Площадь всего многоугольника равна числу всех наблюдаемых явлений, т.е. объему выборки.

Определение. Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длины h, а высоты - $\frac{h_S}{h}$ (плотность частоты). Площадь гистограммы частот равна сумме выборки.

Пример.

Значение x_i	-2	0	1	2	3	5	7
Частота n_i	4	5	7	8	6	2	1

$\Delta l = 7 - (-2) = 9, \quad \Delta l_S = h = \frac{9}{4} = 2{,}25$.

Проме-жуток	[-2;0,25]	[0,25;2,5]	[2,5;4,75]	[4,75;7]
n_S	9	15	6	3
$\frac{n_S}{h}$	4	$\frac{20}{3}$	$\frac{8}{3}$	$\frac{4}{3}$

По гистограмме уже можно себе представить плотность вероятности случайной величины, а по эмпирическим функциям распределения – приближенную теоретическую функцию распределения.