Элементы математической статистики

Краткая историческая справка

Математическая статистика возникла (XVII в.) и развивалась параллельно с теорией вероятностей. Дальнейшее развитие математической статистики (вторая половина XIX—начало XX в.) обязано, в первую очередь, П. Л. Чебышеву, А. А. Маркову, А. М. Ляпунову, а также К. Гауссу, А. Кетле, Ф. Гальтону, К. Пирсону и др.

В XX в. наиболее существенный вклад в математическую статистику был сделан советскими математиками (В. И. Романовский, Е. Е. Слуцкий, А. Н. Колмогоров, Н. В. Смирнов), а также английскими (Стьюдент, Р. Фишер, Э. Пирсон) и американскими (Ю. Нейман, А. Вальд) учеными.

Установление закономерностей, которым подчинены массовые случайные явления, основано на изучении методами теории вероятностей статистических данных — результатов наблюдений.

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:

а) 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.;

б) 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.

Итак, задача математической статистики состоит в создании методов сбора и обработки статистических данных для получения научных и практических выводов.

Выборочный метод

Генеральная и выборочная совокупности

Пусть требуется изучить совокупность однородных объектов относительно некоторого *качественного* или *количественного признака*, характеризующего эти объекты. Например, если имеется партия деталей, то качественным признаком может служить стандартность детали, а количественным — контролируемый размер детали.

Иногда проводят сплошное обследование, т. е. 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.

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 $n = 100$.

Повторная и бесповторная выборки. Репрезентативная выборка

При составлении выборки можно поступать двумя способами: после того как объект отобран и над ним произведено наблюдение, он может быть возвращен либо не возвращен в генеральную совокупность. В соответствии со сказанным выборки подразделяют на повторные и бесповторные.

Повторной называют выборку, при которой отобранный объект (перед отбором следующего) возвращается в генеральную совокупность.

Бесповторной называют выборку, при которой отобранный объект в генеральную совокупность не возвращается.

На практике обычно пользуются бесповторным случайным отбором.

Для того чтобы по данным выборки можно было достаточно уверенно судить об интересующем признаке генеральной совокупности, необходимо, чтобы объекты выборки правильно его представляли. Другими словами, выборка должна правильно представлять пропорции генеральной совокупности. Это требование коротко формулируют так: выборка должна быть *репрезентативной* (*представительной*).

В силу закона больших чисел можно утверждать, что выборка будет репрезентативной, если ее осуществить случайно: каждый объект выборки отобран случайно из генеральной совокупности, если все объекты имеют одинаковую вероятность попасть в выборку.

Если объем генеральной совокупности достаточно велик, а выборка составляет лишь незначительную часть этой совокупности, то различие между повторной и бесповторной выборками стирается; в предельном случае, когда рассматривается бесконечная генеральная совокупность, а выборка имеет конечный объем, это различие исчезает.

Способы отбора

На практике применяются различные способы отбора. Принципиально эти способы можно подразделить на два вида:

1. Отбор, не требующий расчленения генеральной совокупности на части. Сюда относятся: а) простой случайный бесповторный отбор; б) простой случайный повторный отбор.

2. Отбор, при котором генеральная совокупность разбивается на части. Сюда относятся: а) типический отбор; б) механический отбор; в) серийный отбор.

Простым случайным называют такой отбор, при котором объекты извлекают по одному из всей генеральной совокупности. Осуществить простой отбор можно различными способами. Например, для извлечения n объектов из генеральной совокупности объема N поступают так: выписывают номера от 1 до N на карточках, которые тщательно перемешивают, и наугад вынимают одну карточку; объект, имеющий одинаковый номер с извлеченной карточкой, подвергают обследованию; затем карточку возвращают в пачку и процесс повторяют, т. е. карточки перемешивают, наугад вынимают одну из них и т. д. Так поступают n раз; в итоге получают простую случайную повторную выборку объема n.

Если извлеченные карточки не возвращать в пачку, то выборка является простой случайной бесповторной.

Типическим называют отбор, при котором объекты отбираются не из всей генеральной совокупности, а из каждой ее «типической» части. Например, если детали изготовляют на нескольких станках, то отбор производят не из всей совокупности деталей, произведенных всеми станками, а из продукции каждого станка в отдельности. Типическим отбором пользуются тогда, когда обследуемый признак заметно колеблется в различных типических частях генеральной совокупности. Например, если продукция изготовляется на нескольких машинах, среди которых есть более и менее изношенные, то здесь типический отбор целесообразен.

Механическим называют отбор, при котором генеральную совокупность «механически» делят на столько групп, сколько объектов должно войти в выборку, а из каждой группы отбирают один объект. Например, если нужно отобрать 20% изготовленных станком деталей, то отбирают каждую пятую деталь; если требуется отобрать 5% деталей, то отбирают каждую двадцатую деталь, и т. д.

Серийным называют отбор, при котором объекты отбирают из генеральной совокупности не по одному, а «сериями», которые подвергаются сплошному обследованию. Например, если изделия изготовляются большой группой станков-автоматов, то подвергают сплошному обследованию продукцию только нескольких станков. Серийным отбором пользуются тогда, когда обследуемый признак колеблется в различных сериях незначительно.

На практике часто применяют комбинированные методы отбора

Статистическое распределение выборки

Пусть из генеральной совокупности извлечена выборка, причем x_1 наблюдалось n_1 раз, $x_2 — n_2$ раз, $x_k — n_k$ раз и $\sum n_i = n$ — 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. Статистическое распределение можно задать также в виде последовательности интервалов и соответствующих им частот (в качестве частоты, соответствующей интервалу, принимают сумму частот, попавших в этот интервал).

Заметим, что в теории вероятностей под *распределением* понимают соответствие между возможными значениями случайной величины и их вероятностями, а в математической статистике — соответствие между наблюдаемыми вариантами и их частотами, или относительными частотами.

Пример. Задано распределение частот выборки объема $n = 20$:

x_i	2	6	12
n_i	3	10	7

Написать распределение относительных частот.

Решение. Найдем относительные частоты, для чего разделим частоты на объем выборки:

$$W_1 = 3/20 = 0,15, \quad W_2 = 10/20 = 0,50, \quad W_3 = 7/20 = 0,35.$$

Напишем распределение относительных частот:

x_i	2	6	12
W_i	0,15	0,50	0,35

Контроль: $0,15 + 0,50 + 0,35 = 1$.

Полигон и гистограмма

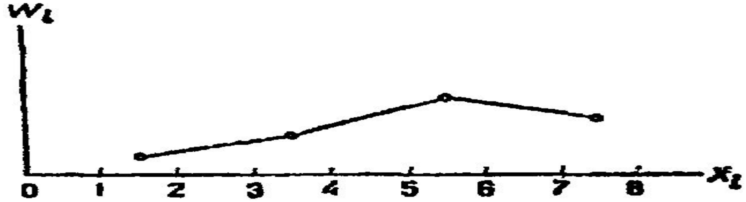
Для наглядности строят различные графики статистического распределения и, в частности, полигон и гистограмму.

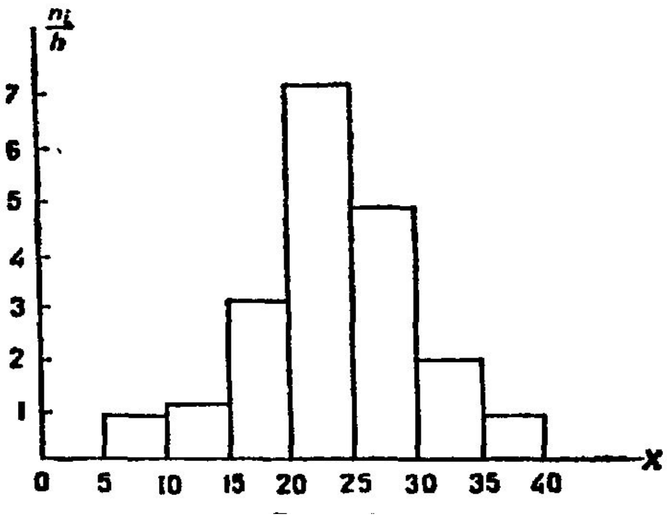
Полигоном частот называют ломаную, отрезки которой соединяют точки $(x_1; n_1)$, $(x_2; n_2)$, ..., $(x_k; n_k)$. Для построения полигона частот на оси абсцисс откладывают варианты x_i, а на оси ординат — соответствующие им частоты n_i. Точки $(x_i; n_i)$ соединяют отрезками прямых и получают полигон частот.

Полигоном относительных частот называют ломаную, отрезки которой соединяют точки $(x_1; W_1)$, $(x_2; W_2)$, ...

..., $(x_k; W_k)$. 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





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

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

Для построения гистограммы относительных частот на оси абсцисс откладывают частичные интервалы, а над ними проводят отрезки, параллельные оси абсцисс на расстоянии W_i/h. Площадь i-го частичного прямоугольника равна $hW_i/h = W_i$ — относительной частоте вариант, попавших в i-й интервал. Следовательно, *площадь гистограммы относительных частот равна сумме всех относительных частот, т. е. единице.*