Лекция. Понятие озадачах математической статистики

1. Задачи математической статистики[1]. Установление закономерностей, которым подчинены массовые случайные явления, основано на изучении статистических данных — результатах наблюдений. *Первая задача* математической статистики — указать способы сбора и группировки (если данных очень много) статистических сведений.

Вторая задача математической статистики — разработать методы анализа статистических данных в зависимости от целей исследования.

Изучение тех или иных явлений методами математической статистики служит средством решения многих вопросов, выдвигаемых наукой и практикой (правильная организация технологического процесса, наиболее целесообразное планирование и др.).

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

2. 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 применяется 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.

---

[1] Термин «статистика» происходит от латинского слова *status* — состояние.

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 $n = 100$.

Замечание. Часто генеральная совокупность содержит конечное число объектов. Однако если это число достаточно велико, то иногда в целях упрощения вычислений или для облегчения теоретических выводов допускают, что генеральная совокупность состоит из бесчисленного множества объектов. Такое допущение оправдывается тем, что увеличение объема генеральной совокупности (достаточно большого объема) практически не сказывается на результатах обработки данных выборки.

3. Выборка с возвращением и без возвращения. Репрезентативная выборка. При составлении выборки можно поступать двояко: после того как объект отобран и над ним произведено наблюдение, он может быть либо возвращен, либо не возвращен в генеральную совокупность. В соответствии со сказанным выборки подразделяют на выборки с возвращением и без возвращения.

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

Выборкой без возвращения называют выборку, при которой отобранный объект в генеральную совокупность не возвращается.

На практике обычно пользуются выборкой без возвращения.

Для того чтобы по данным выборки можно было достаточно уверенно судить об интересующем нас признаке генеральной совокупности, необходимо, чтобы объекты выборки правильно его представляли. 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 выборкой с возвращением и без возвращения стирается; в предельном случае, когда рассматривается бесконечная генеральная совокупность, а выборка имеет конечный объем, это различие исчезает.

4. 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, к нему относятся:

а) простой случайный бесповторный отбор;

б) простой случайный повторный отбор.

2. Отбор, при котором генеральная совокупность разбивается на части, к нему 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 объемом 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 карточка возвращается в пачку и процесс повторяется, т. е. карточки перемешиваются, наугад вынимают одну из них и т. д. Так поступают n раз; в итоге получают простую случайную выборку с возвращением объемом n.

Если извлеченные карточки не возвращать в пачку, то выборка будет простой случайной без возвращения.

При большом объеме генеральной совокупности описанный процесс оказывается очень трудоемким. В этом случае пользуются готовыми таблицами случайных чисел, в которых числа расположены в случайном порядке. Для того чтобы отобрать, например, 50 объектов из пронумерованной генеральной совокупности, открывают любую страницу таблицы случайных чисел и выписывают подряд 50 чисел; в выборку попадают те объекты, номера которых совпадают с выписанными случайными числами. Если окажется, что случайное число таблицы превышает число N, то такое случайное число пропускают. При осуществлении выборки без возвращения случайные числа таблицы, уже встречавшиеся ранее, следует также пропустить.

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 генеральная совокупность «механически» делится на столько групп, сколько объектов должно войти в выборку, и из каждой группы отбирается 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.

Следует указать, что иногда механический отбор может не обеспечить репрезентативности выборки. Например, если отбирается каждый двадцатый обтачиваемый валик, причем сразу же после отбора производят замену резца, то отобранными окажутся все валики, обточенные затупленными резцами. В таком случае надо устранить совпадение ритма отбора с ритмом замены резца, для чего надо отбирать, скажем, каждый десятый валик из двадцати обточенных.

Серийным называют отбор, при котором объекты отбирают из генеральной совокупности не по одному, а «сериями», подвергающимися 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.

Подчеркнем, что на практике часто применяется комбинированный отбор, при котором сочетаются указанные выше способы.

Например, иногда разбивают генеральную совокупность на серии одинакового объема, затем простым случайным отбором выбирают несколько серий и, наконец, из каждой серии простым случайным отбором извлекают отдельные объекты.

Рассмотрим несколько задач.

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 17.1

Составить распределение относительных частот, если задано распределение частот выборки объемом $n = 20$:

x_i	2	6	12
n_i	3	10	7

Решение

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

$$W_1 = \frac{3}{20} = 0{,}15;\ W_2 = \frac{10}{20} = 0{,}5;\ W_3 = \frac{7}{20} = 0{,}35.$$

Составим распределение относительных частот:

x_i	2	6	12
W_i	0,15	0,5	0,35

Сумма относительных частот составляет 0,15 + 0,5 + 0,35 = 1.

---

2. Эмпирическая функция распределения. Пусть известно статистическое распределение частот количественного признака X. Введем обозначения: n_x — число наблюдений, при которых наблюдалось значение признака меньше x, n — общее число наблюдений (объем выборки).

Относительная частота события $X < x$ равна n_x/n. Если x будет изменяться, то, вообще говоря, будет изменяться и относительная частота, т. е. относительная частота n_x/n есть функция от x. Так как эта функция находится эмпирическим (опытным) путем, то ее называют эмпирической.

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

$$F^*(x) = \frac{n_x}{n},$$

где n_x — число вариант, меньших x; n — объем выборки.

Таким образом, для того чтобы найти, например, $F^*(x_2)$, надо число вариант, меньших x_2, разделить на объем выборки:

$$F^*(x_2) = \frac{n_{x_2}}{n}.$$

Свойства

1) значения эмпирической функции принадлежат отрезку [0; 1];

2) $F^*(x)$ — неубывающая функция;

3) если x_1 — наименьшая варианта, то $F^*(x) = 0$ при $x \le x_1$; если x_k — наибольшая варианта, то $F^*(x) = 1$ при $x > x_k$.

Итак, эмпирическая функция распределения выборки служит для оценки теоретической функции распределения генеральной совокупности.

Пример 17.2

Построить эмпирическую функцию по данному распределению выборки:

Варианты x_i	2	6	10
Частоты n_i	12	18	30

Решение. Найдем объем выборки: $12 + 18 + 30 = 60$. Наименьшая варианта равна 2, следовательно, при $x \le 2$ $F^*(x) = 0$.

Значение $X < 6$, а именно $x_1 = 2$, наблюдалось 12 раз; следовательно, при $2 < x \le 6$

$$F^*(x) = \frac{12}{60} = 0{,}2.$$

Значения $X < 10$, а именно $x_1 = 2$ и $x_2 = 6$, наблюдались $12 + 18 = 30$ раз; следовательно, при $6 < x \le 10$

$$F^*(x) = \frac{30}{60} = 0{,}5.$$

Так как $x = 10$ — наибольшая варианта, то при $x > 10$ $F^*(x) = 1$.

Искомая эмпирическая функция

$$F^*(x) = \begin{cases} 0 \text{ при } x \le 2, \\ 0{,}2 \text{ при } 2 < x \le 6, \\ 0{,}5 \text{ при } 6 < x \le 10, \\ 1 \text{ при } x > 10. \end{cases}$$

График этой функции изображен на рис. 17.1.

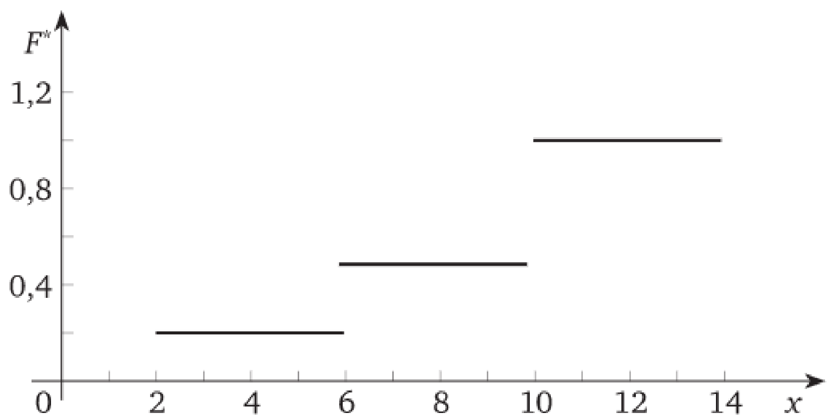

Рис. 17.1

3. Полигон и гистограмма. В целях 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.

На рис. 17.2 изображен полигон относительных частот следующего распределения:

X	1,5	3,5	5,5	7,5
W	0,1	0,2	0,4	0,3

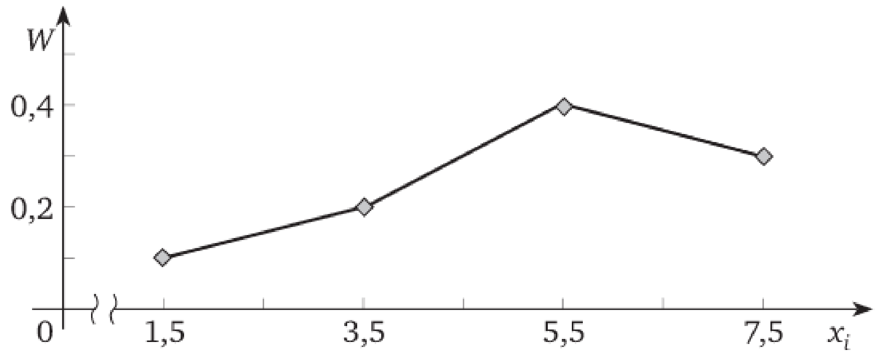

Рис. 17.2

В случае непрерывного признака целесообразно строить гистограмму, для чего интервал, в котором заключены все наблюдаемые значения признака, разбивают на несколько частичных интервалов длиной h и находят для каждого частичного интервала n_i — сумму частот вариант, попавших в i-й интервал.

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

Для построения гистограммы частот на оси абсцисс откладывают частичные интервалы, а над ними проводят отрезки, параллельные оси абсцисс, на расстоянии n_i/h.

Площадь i-го прямоугольника равна $h \cdot \frac{n_i}{h} = n_i$ — сумме частот вариант i-го интервала; следовательно, *площадь гистограммы частот равна сумме всех частот, т. е. объему выборки.*

На рис. 17.3 изображена гистограмма частот распределения объема $n = 100$, приведенного в таблице.

Частичный интервал длиной $h = 5$	**Сумма частот вариант частичного интервала n_i**	**Плотность частоты $\frac{n_i}{h}$**
5—10	4	0,8
10—15	6	1,2
15—20	16	3,2
20—25	36	7,2
25—30	24	4,8
30—35	10	2,0
35—40	4	0,8

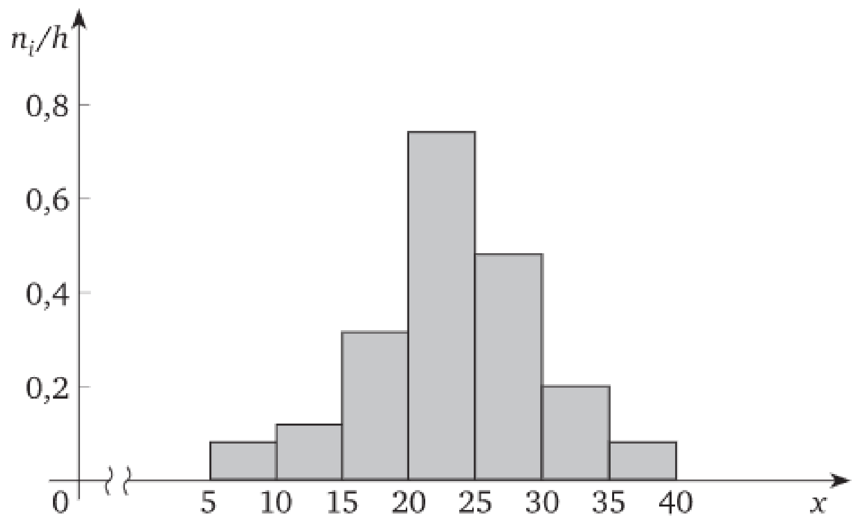

Рис. 17.3

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

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

Выводы:

Математическая статистика – это наука, изучающая методы сбора и обработки статистической информации для получения научных и практических выводов.

Основным методом матстатистики является **выборочный метод**, его суть состоит в исследовании представительной *выборочной совокупности* – для достоверной характеристики совокупности генеральной. Данный метод экономит временные, трудовые и материальные затраты, поскольку исследование всей совокупности зачастую затруднено или невозможно.

Глава 11 «Элементы теории вероятности и математической статистики», учебник Башмаков М.И. Математика: алгебра и начала математического анализа, геометрия: учеб. для студ. учреждений сред.проф. образования/ М.И. Башмаков. – 4-е изд.,стер. – М. : ИЦ «Академия», 2017, - 256 с.

В случае отсутствия печатного издания, Вы можете обратиться к Электронно-библиотечной системе.

Список использованных интернет-ресурсов:

1. https://23.edu-reg.ru/
2. https://urait.ru/