

А. З. Абилгасанова, Н. А. Нечаева
Уральский Государственный Экономический Университет
г. Екатеринбург

Тема: «Математическая статистика.»

Математическая статистика — наука, разрабатывающая математические методы систематизации и использования статистических данных для выводов как научных, так и практических.

Математическая статистика во многих своих разделах основывается на теорию вероятностей, дающая возможность оценить надёжность и точность выводов, делаемых на основании ограниченного статистического материала. Например, оценить необходимый объём выборки для получения результатов требуемой точности при выборочном обследовании.

Предмет и методы.

Математическая статистика — раздел математики, разрабатывающий методы регистрации, описания и анализа данных наблюдений и экспериментов. Его целью является построения вероятностных моделей массовых случайных явлений. Математическая статистика делится на статистику чисел, многомерный статистический анализ, анализ функций (процессов) и временных рядов, статистику объектов нечисловой природы.

Выделяют несколько методов статистики: описательную, теорию оценивания и теорию проверки гипотез.

Описательная статистика представляет собой совокупность эмпирических методов, которые используют для визуализации и интерпретации данных (расчет выборочных характеристик, таблицы, диаграммы, графики и т. д.), как правило, не требующих предположений о вероятностной природе данных. Некоторые методы описательной статистики предполагают использование возможностей современных компьютеров. К ним относятся, к примеру, кластерный анализ, нацеленный на выделение групп объектов, похожих друг на друга, и многомерное шкалирование, которое позволит наглядно представить объекты на плоскости.

Методы оценивания и проверки гипотез опираются на вероятностные модели происхождения данных. Эти модели делятся на параметрические и непараметрические. Параметрические модели предполагают, что

характеристики изучаемых объектов описываются посредством распределений, зависящих от одного или нескольких числовых параметров. Непараметрические модели не связаны со спецификацией параметрического семейства для распределения изучаемых характеристик. В математической статистике оценивают параметры и функции от них, представляющие важные характеристики распределений (например, математическое ожидание, медиана, стандартное отклонение, квантили и др.), плотности и функции распределения и прочее. Используют точечные и интервальные оценки.

Огромный раздел современной математической статистики — статистический последовательный анализ, фундаментальный вклад в создание и развитие которого внес А. Вальд во время Второй мировой войны. В отличие от непоследовательных (традиционных) методов статистического анализа, которые основаны на случайной выборке фиксированного объема, в последовательном анализе допускается формирование массива наблюдений по одному (или, более общим образом, группами), при этом решение об проведении следующего наблюдения (группы наблюдений) принимается на основе уже накопленного массива наблюдений. Исходя из этого, теория оптимальной остановки тесно связана с теорией последовательного статистического анализа.

В математической статистике имеется общая теория проверки гипотез и масса методов, которые посвящены проверке конкретных гипотез. Рассматривают гипотезы о значениях параметров и характеристик, о проверке однородности (то есть о совпадении характеристик или функций распределения в двух выборках), о согласии эмпирической функции распределения с заданной функцией распределения или с параметрическим семейством таких функций, о симметрии распределения и др.

Огромное значение имеет раздел математической статистики, который связан со свойствами различных схем организации выборок, с проведением выборочных обследований и построением адекватных методов оценивания и проверки гипотез.

Задачи восстановления зависимостей активно изучаются более 200 лет, с момента разработки К. Гауссом в 1794 г. метода наименьших квадратов.

Когда Карл Пирсон создал метод главных компонент, тогда было начато создание методов аппроксимации данных и сокращения размерности описания. В дальнейшем были разработаны факторный анализ и многочисленные нелинейные обобщения.

Разного рода методы построения (кластер-анализ), анализа и использования (дискриминантный анализ) классификаций (типологий) именуют также

методами распознавания образов (с учителем и без), автоматической классификации и др.

На данный момент времени в математической статистике огромную роль играют компьютеры. Их используют как для расчётов, так и для имитационного моделирования (в частности, в методах размножения выборок и при изучении пригодности асимптотических результатов).

Связь математической статистики с теорией вероятностей.

Связь между математической статистики и теорией вероятностей имеет в разных случаях различный характер. Теория вероятностей изучает не все явления, а те явления, которые случайные и именно «вероятностно случайные», т. е. такие, для которых имеет смысл говорить о соответствующих им распределениях вероятностей. Так или иначе теория вероятностей играет некоторую определённую роль и при статистическом изучении многочисленных явлений любой природы, которые могут не относиться к категории вероятностно случайных. Это осуществляется через основанные на теории вероятностей теорию выборочного метода и теорию ошибок измерений. В этих случаях вероятностным закономерностям подчинены не сами изучаемые явления, а приёмы их исследования.

При статистическом исследовании вероятностных явлений в полной мере находят применение такие основанные на теории вероятностей разделы математической статистики, как теория проверки статистических гипотез, теория статистические оценки распределений вероятностей или параметров этих распределений. При использовании таких разделов математической статистике необходимо, чтобы сами изучаемые явления подчинялись достаточно определённым вероятностным закономерностям. Например, статистическое изучение режима турбулентных водных потоков или флуктуаций в радиоприёмных устройствах производится на основе теории стационарных случайных процессов. Однако применение той же теории к анализу случайных процессов в экономике может привести к грубым ошибкам ввиду того, что входящее в определение стационарного процесса предположение о наличии сохраняющихся в течение длительного времени неизменных распределений вероятностей в этом случае, как правило, неприемлемо.

Вероятностные закономерности проявляются в статистических данных в силу больших чисел закона (например, частоты событий близки к их вероятностям, а средние значения – к математическим ожиданиям).

Задача

Измерение роста детей младшей группы детского сада представлено выборкой:

92, 96, 95, 96, 94, 97, 98, 94, 95, 96.

Найдем некоторые характеристики этой выборки.

Решение

Размер выборки (число измерений; N): 10.

Наименьшее значение выборки: 92. Наибольшее значение выборки: 98.

Размах выборки: $98 - 92 = 6$.

Запишем ранжированный ряд (варианты в порядке возрастания):

92, 94, 94, 95, 95, 96, 96, 96, 97, 98.

Сгруппируем ряд и запишем в таблицу (каждой варианте поставим в соответствие число ее появлений):

x_i	92	94	95	96	97	98	N
n_i	1	2	2	3	1	1	10

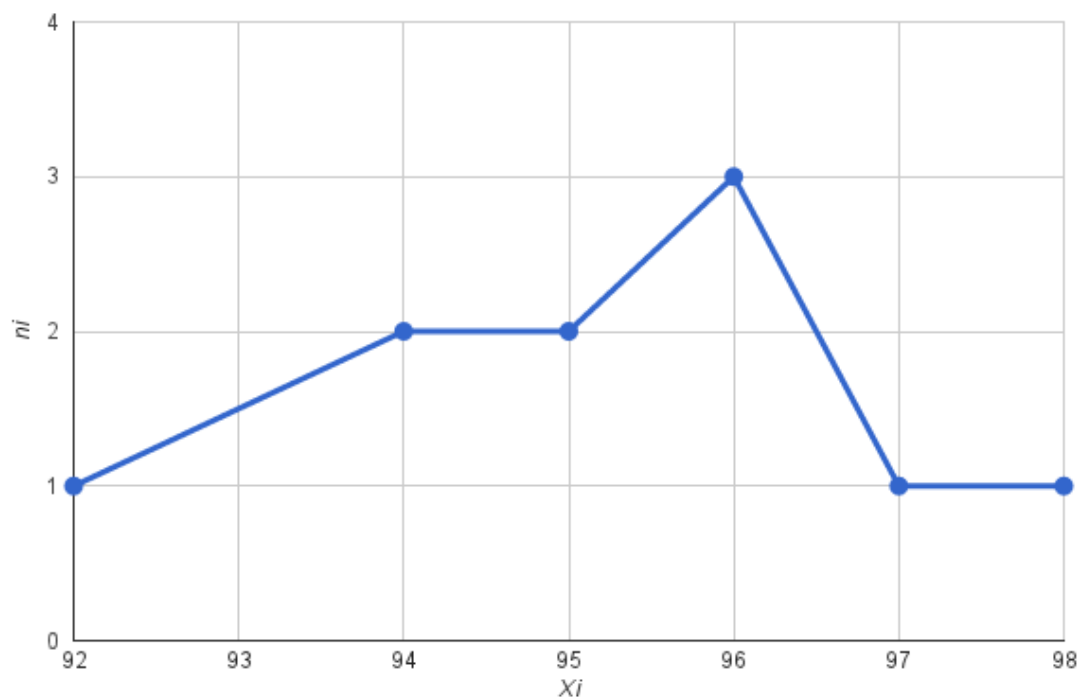
Вычислим относительные частоты и накопленные частоты, результат запишем в таблицу:

Чтобы определить накопительную частоту следует взять последнюю накопительную частоту и прибавить к ней абсолютную частоту данной величины.

x_i	92	94	95	96	97	98	N
n_i	1	2	2	3	1	1	10
$w_i = \frac{n_i}{n}$	0,1	0,2	0,2	0,3	0,1	0,1	1

Накопленные частоты	1	3	5	8	9	10
---------------------	---	---	---	---	---	----

Построим полигон частот выборки (отметим на графике варианты по оси Ox , частоты по оси Oy , соединим точки линией).



Выборочную среднюю и дисперсию вычислим по формулам (соответственно):

$$\bar{x} = \frac{1}{N} \sum_{i=1}^n x_i n_i; \quad D = \frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})^2 n_i.$$

$$\bar{x} = \frac{1}{10} \cdot (92 \cdot 1 + 94 \cdot 2 + 95 \cdot 2 + 96 \cdot 3 + 97 \cdot 1 + 98 \cdot 1) = 95,3$$

$$D = \frac{1}{10} \cdot ((92 - 95,3)^2 \cdot 1 + (94 - 95,3)^2 \cdot 2 + (95 - 95,3)^2 \cdot 2 + (96 - 95,3)^2 \cdot 3 + (97 - 95,3)^2 \cdot 1 + (98 - 95,3)^2 \cdot 1) = 2,61$$

Можно находить и другие характеристики выборки, но для общего представления вполне достаточно найденных характеристик.

В заключении отметим, математическая статистика разрабатывает методы описания и анализа данных наблюдений и экспериментов, цель которого построить вероятностные модели массовых случайных явлений. Связь между математической статистики и теорией вероятностей имеет в разных случаях различный характер.

Руководитель: Синцова С.Г.

Библиографический список:

1. Высшая математика. Математический анализ: учеб. пособие / Ю.Б. Мельников, М.Д. Боярский, М.Д. Локшин, П.И. Гниломедов, С.Г. Синцова, А.А. Кныш; М-во науки и высш. образования Рос. Федерации,

Урал. Гос. экон. ун-т. – Екатеринбург: Изд-во Урал. гос. экон. ун-та,
2018

2. Ван дер Варден Б. Л. Математическая статистика. М., 1960
3. Ивченко Г. И., Медведев Ю. И. Математическая статистика. 2-е изд. М., 1992
4. Боровков А. А. Математическая статистика. 4-е изд. СПб., 2010.