**Тема: «Понятие о задачах математической статистики»**

**Математическая статистика** – это раздел математики который занимается разработкой методов сбора, описания и анализа экспериментальных результатов наблюдений, массовых случайных явлений.

**Статистические характеристики** – это математические понятия, с помощью которых описываются отличительные особенности и свойства совокупности данных, полученных с помощью наблюдений или каким-то другим способом.

**Значение характеристик** состоит еще и в том, что они «**подсказывают», с каких позиций целесообразно анализировать** имеющуюся совокупность данных.

**Сбор информации** происходит массово или выборочно. При этом используется: перепись населения, отчеты предприятий, текущий учет, опросы, анкетирование, интервьюирование, наблюдения, статистика больниц, загсов и т.д.

**Фундаментальными понятиями математической статистики являются *генеральная совокупность и выборка.*** Генеральную совокупность удобно изображать с использованием круговой диаграммы, выборку – с использованием части круговой диаграммы.

**Способы образования выборочной совокупности:** случайная (отбирая на удачу), механическая (отбирая через определенный интервал), типическая (случайные выборки из каждой группы), серийная (разбивается на непересекающиеся серии или группы).

**Обработка собранной информации.**

Статистическая информация о результатах наблюдений или экспериментов может быть представлена в различных формах.

Простейшей из них является запись в порядке их появления – запись в ряд:

$x\_{1, }x\_{2, }x\_{3, }$…$x\_{n}$, называемый ***простым статистическим рядом*** или ***выборкой.***

Отдельные значения $x\_{i}$, составляющие этот ряд, называют ***вариантами*** или просто данными.

**Понятие объема ряда**

Количество вариант (данных) в ряду ***n*** называют ***объемом ряда***, или объемом выборки.

Варианты в ряду могут иметь как различные, так и одинаковые значения.

**Понятие ранжированного ряда**

Составить ***ранжированный ряд*** - это значит записать варианты в порядке их возрастания.

**Характеристики числового ряда**

*Пример 1. Пусть ученик получил в течение года следующие отметки по математике:5, 2, 4, 5, 5, 4, 4, 5, 5, 5. Какую четвертную отметку поставит ему учитель?*

Многих школьников волнует подобная проблема, и чаще всего ученики решают ее следующим естественным образом: складывают все отметки и делят сумму оценок на их количество.

В нашем случае*(5 + 2 + 4 + 4 + 5 + 5 + 4 + 4 + 5 + 5 + 5) / 10 = 4,4*

Число 4,4, которое получается в результате, называется **средним арифметическим**. Поскольку такую оценку в журнал ставить не принято, учитель, скорее всего, округлит ее до 4.

**Средним арифметическим (или выборочным средним)** ряда чисел называется частное от деления суммы этих чисел на их количество:

$\overbar{x}$ **=** $\frac{x\_{1 }+x\_{2 }+x\_{3 }+…+x\_{n. }}{n}$

Среднее арифметическое, конечно, является важной характеристикой ряда чисел, в нашем случае — отметок за четверть, но иногда полезно рассматривать и другие средние.

Например, претендуя на «5», ученик наверняка будет использовать такой аргумент: «Чаще всего в четверти я получал пятерки!». Статистик в этом случае сказал бы иначе: «Модой этого ряда является число 5».

**Модой** (Мо) называют число ряда, которое встречается в этом ряду наиболее часто.

Можно сказать, что оно в этом ряду самое «модное». В отличие от среднего арифметического, которое можно вычислить для любого числового ряда, моды может вообще не быть.

Например, пусть тот же ученик получил по русскому языку следующие отметки: 4, 2, 3, 5. Каждая отметка встречается в этом ряду только один раз, и среди них нет числа, встречающегося чаще других. Значит, у этого ряда нет моды. А вот среднее арифметическое, конечно, есть:*(4 + 2 + 3 + 5) : 4 = 3,5*.

Такой показатель, как мода, можно использовать не только в числовых рядах. Вы уже знакомы с социологическими опросами. Если, например, опросить большую группу учеников, какой предмет вам нравится больше всего, то модой можно назвать тот предмет, который будут называть чаще остальных. Это одна из причин, по которой мода широко используется при изучении спроса. Например, при решении вопросов, в пачки какого веса фасовать масло, какие открывать авиарейсы и т. п., предварительно изучается спрос и выявляется мода — наиболее часто встречающийся заказ. И даже выборы президента, с точки зрения статистики, не более чем определение моды.

**Медиана числового ряда**

***Медианой ряда, состоящего из нечетного количества чисел***, называется число данного ряда, которое окажется посередине, если этот ряд упорядочить: $x\_{1}$**,**$x\_{2},x\_{3},x\_{4,}x\_{5}$

**Me =** $x\_{3}$

***Медианой ряда, состоящего из четного количества чисел***, называется среднее арифметическое двух стоящих посередине чисел этого ряда, если этот ряд упорядочить.$: x\_{1}$**,**$x\_{2},x\_{3},x\_{4,}x\_{5},x\_{6}$

**Me =** $\frac{x\_{3}+x\_{4,}}{2}$

Для того чтобы найти **медиану** ряда чисел, нужно сначала их упорядочить — составит **ранжированный ряд (**записать в порядке возрастания).

*Пример 2.В конце года 11 учеников 8 класса сдавали норматив по бегу на 100 метров. Были зафиксированы следующие результаты:*

|  |  |
| --- | --- |
| Ученик | Результат(с) |
| Данила | 15,3 |
| Петя | 16,9 |
| Лена | 21,8 |
| Катя | 18,4 |
| Стаc | 16,1 |
| Аня | 25,1 |
| Оля | 19,9 |
| Боря | 15,5 |
| Паша | 14,7 |
| Наташа | 20,2 |
| Миша | 15,4 |

*После того как все ребята пробежали дистанцию, к преподавателю подошел Петя и спросил, какой у него результат.* «Самый средний результат: 16,9 секунды», — ответил учитель. «Почему? — удивился Петя. — Ведь среднее арифметическое всех результатов — примерно 18,3 секунды, а я пробежал на секунду с лишним лучше. И вообще, результат Кати (18,4) гораздо ближе к среднему, чем мой». «Твой результат средний, потому что пять человек пробежали лучше, чем ты, и пять — хуже. То есть ты как раз посередине», — сказал учитель. На языке статистики результат Пети называется *медианой* исходного ряда данных.

 Для того чтобы найти медиану ряда чисел, нужно сначала их упорядочить — составить ранжированный ряд. В нашем примере он выглядит так:

14,7; 15,3; 15,4; 15,5; 16,1; **16,9**; 18,4; 19,9; 20,2; 21,8; 25,1.

Средним (шестым по счету) числом является 16,9: пять чисел меньше него, пять чисел больше. Значит, 16,9 — медиана.

Достоинством медианы является ее большая по сравнению со средним арифметическим **«устойчивость к ошибкам».**

Представим себе, что в наши наблюдения вкралась досадная оплошность: например, при записи одного из результатов соревнований мы пропустили десятичную запятую и вместо 20,2 написали 202. Тогда среднее арифметическое результатов возрастет с 18,1 секунды до 34,6 секунды, а медиана будет по-прежнему 16,9 секунды!

**Медиану используют вместо средней арифметической, когда крайние варианты упорядоченного ряда (наименьшая и наибольшая) по сравнению с остальными оказываются чрезмерно большими или чрезмерно малыми.**

Числовой ряд иногда удобно представлять в виде таблицы, если имеем большой объем информации и данные повторяются.

Представим ряд данных *5, 2, 4, 5, 5, 4, 4, 5, 5, 5* в виде таблицы

|  |  |  |  |
| --- | --- | --- | --- |
| X | 2 | 4 | 5 |
| M | 1 | 3 | 6 |
| W | 1/1010$\%$ | 3/1030$\%$ | 6/1060$\%$ |

В первой строке – значение случайной величины Х, во второй – частота значений варианты М, в третьей строке – относительная частота появления события.

По табличным данным тоже можно найти объем ряда, среднее арифметическое, моду и медиану.

Объем - сложить все данные М

Мода – самое большое значение М

Медиана - Ме$ \geq 0,5W$

Среднее арифметическое: $\overbar{x}$=cумме произведений элементов первой строки на частоту появления второй строки, и все поделить на 10:$\frac{2∙1+4∙3+5∙6}{10}$

**Статистические характеристики: среднее арифметическое, мода, медиана называются средними результатами измерений.**

Обработанные результаты статистики можно демонстрировать графически.

*Пример 3: В первом полугодии 2011 года завод получил прибыль в 10 млн. рублей. Распределение прибыли по месяцам показано в таблице*

****

В координатной плоскости на оси абсцисс будем отмечать номер месяца (янв. – 1, февр. – 2 и т.д.). На оси ординат будем отмечать прибыль завода
 (в млн. руб.).
Отметим точки: (1;1,4),(2;1,3),(3;1,5),(4;2,1),(5;2),(6;1,7) и соединим их последовательно отрезками.

Полученную ломаную линию называют ***полигоном частот.***



**Графики статистического распределения.**

Распределение случайных величин можно задавать и демонстрировать графически.



**Гистограмма –** помогает наглядно сравнивать по величине несколько объектов.



**Полигон частот –** показывает промежутки убывания и возрастания, точки максимума и минимума.



**Круговая диаграмма.** Круговые диаграммы используют в тех случаях, когда нужно показать части какого-либо целого.

Существуют и другие статистические характеристики, которые называются характеристиками отклонения.

**Размах. Дисперсия. Среднеквадратичное отклонение**

Средние характеристики числового ряда позволяют оценить его поведение «в среднем». Но это далеко не всегда полностью характеризует выборку.

**Размах** — это разность наибольшего и наименьшего значений ряда данных: $x\_{1, }x\_{2, }x\_{3, }$**…**$x\_{n}$

**R =** $x\_{n}- x\_{1}$

*Пример 4.Температура на Меркурии колеблется от - 150*$°$ *до + 350*$°.$ *Удобен ли климат Меркурия для жизни людей, если на планете Меркурий средняя температура +15*$°$*?*

**Например,** на планете **Меркурий с**редняя температура +15°. Исходя из этого статистического показателя, можно подумать, что на Меркурии умеренный климат, удобный для жизни людей.

Однако на самом деле это не так. Температура на Меркурии колеблется от — 150° до +350°.

Значит, чтобы получить представление о поведении числового ряда, помимо средних характеристик надо знать **характеристики разброса,** показывающие, насколько значения ряда различаются между собой, как сильно они «разбросаны» вокруг средних. Простейшей такой характеристикой является **размах.**

Для температуры на Меркурии, например, размах равен

350° — (-150°) = 500°. Конечно, такого перепада температур человек выдержать не может.

Размах очень просто вычисляется, но не всегда несет достоверную информацию, так как на его величину может сильно повлиять какое-то одно (возможно, ошибочное) значение статистического ряда.

Вот почему в реальных статистических исследованиях чаще используют другую характеристику разброса, которая сложнее вычисляется, но зато меньше подвержена таким колебаниям.

Прежде чем определять эту величину, рассмотрим на примере, какой самый естественный способ вычисления «среднего отклонения от среднего».

*Пример 5.Дан числовой ряд, который представляет собой стоимость одного литра бензина на 10 автозаправочных станциях (в рублях):*

*32,2; 32,8; 33; 32,9; 33; 32,5; 32,8; 33; 33,2; 32,8.*

Найдем среднее арифметическое этих цен:

*(32,2 + 32,8 + 33 + 32,9 + 33 + 32,5 + 32,8 + 33+ 33,2 + 32,8 ) / 10 = 32,82.*

Самым естественным, на первый взгляд, кажется посчитать отклонение от среднего для каждого члена ряда и затем найти их среднее арифметическое:

*((32,2 - 32,82) + (32,8 - 32,82) +(33- 32,82) + … + (32,8 - 32,82)) / 10 = 0.*

Мы получили нуль совсем не случайно: при вычислении «среднего разброса» по такой формуле часть отклонений входит в сумму со знаком «плюс», часть — со знаком «минус», а в сумме всегда получается нуль.

Какой же выход? Можно суммировать, например, модули отклонений — тогда уж нуля точно не будет. Иногда так и поступают, но с модулем не всегда удобно работать. Поэтому математики решили, что лучше складывать не модули отклонений, а их квадраты — они ведь тоже неотрицательные.

Так появилось понятие **дисперсии числового ряда**.

**Дисперсией числового ряда** называется среднее арифметическое квадратов отклонений от среднего арифметического.

**D =** $\frac{(x\_{1}- \overbar{x})^{2}+ (x\_{2}- \overbar{x})^{2}+…+(x\_{n}- \overbar{x})^{2}}{n}$

Найдем дисперсию числового ряда из нашего примера с ценами на бензин. Среднее арифметическое мы уже вычислили — оно равно*32,82*.

Найдем теперь **дисперсию**, т. е. **среднее арифметическое квадратов отклонений от среднего:**

*((32,2 - 32,82)2 + (32,8 - 32,82)2 + (33 - 32,82)2 + … + (32,8 - 32,82)2 ) / 10 = 0,0736*.

У дисперсии есть один существенный недостаток: если исходные значения ряда измеряются в каких-то единицах(например, в рублях), то у дисперсии эти единицы возводятся в квадрат («**квадратные»** рубли).

В нашем примере среднее значение цены получилось 32 рубля 82 копейки, а вот дисперсия цен — около 7 … «квадратных копеек».

Избавиться от таких странных единиц измерения можно, если использовать другую характеристику разброса — стандартное отклонение.

**Стандартным (или средним квадратичным) отклонением**числового ряда называется квадратный корень из дисперсии:$σ= \sqrt{D(x)}$

Обозначают его греческой буквой  («сигма»).В рассмотренном примере стандартное отклонение будет $δ= \sqrt{0,0736 }≈0,27$, т.е. приблизительно 27 коп.

Как и при изучении средних характеристик, попробуем найти характеристики разброса **по таблице частот**.

**Математическое ожидание случайной величины**

Как мы знаем, распределение вероятностей случайной величины — это таблица, в которой указаны значения случайной величины и их вероятности. Для практики не всегда нужно изучать всю таблицу распределения. Достаточно знать некоторые ее числовые характеристики. Рассмотрим случайную величину *X*. Ее математическое ожидание обычно обозначают *Е(Х)*.Пусть распределение вероятностей случайной величины *X* задано таблицей:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Значение величины X | http://free.megacampus.ru/xbookM0005/files/31-05f.gif | http://free.megacampus.ru/xbookM0005/files/31-06f.gif | http://free.megacampus.ru/xbookM0005/files/31-07f.gif |   | http://free.megacampus.ru/xbookM0005/files/31-08f.gif |
| Вероятность | http://free.megacampus.ru/xbookM0005/files/31-09f.gif | http://free.megacampus.ru/xbookM0005/files/31-10f.gif | http://free.megacampus.ru/xbookM0005/files/31-11f.gif |   | http://free.megacampus.ru/xbookM0005/files/31-12f.gif |

**Математическим ожиданием** случайной величины *X* называют число

*Е(Х)= х1 · Р1 + х2 · Р2 + … + хn · Рn*.

Математическое ожидание *Е(Х)* называют также **ожидаемым значением** случайной величины *X*, **средним значением** случайной величины *X*. Если значения случайной величины измеряются в каких-либо единицах (например, рост —в сантиметрах, температура —в градусах), то ее математическое ожидание измеряется в этих же единицах (средний рост —в сантиметрах, средняя температура — в градусах)

*Пример 6.Для проведения лотереи изготовили 100 билетов. Из них 1 билет с выигрышем в 500 р., 10 билетов с выигрышами по 100 р. и остальные 89 билетов без выигрышей. Наудачу выбирают один билет. Найдем математическое ожидание выигрышаM(X).*

|  |  |  |  |
| --- | --- | --- | --- |
| Количество билетов | 1 | 10 | 89 |
| Выигрыш | 500 руб. | 100 руб. | 0 руб |
| Вероятность | 1/100 | 1/10 | 89/100 |

Эта случайная величина может принимать три значения: 500 р., 100 р. и 0 р. (нет выигрыша). Их вероятности равны 0,01, 0,10 и 0,89.Математическое ожидание выигрыша равно *500 · 0,01 + 100 · 0,10 + 0 · 0,89 = 15* (р.).Получается, что средний выигрыш на один билет равен 15 р.

Для того чтобы лотерея приносила доход своим устроителям, цена билета должна быть больше, чем средний выигрыш. Предположим, что билет стоит 20 р. Продав все билеты, устроители лотереи получат 2000 рублей.

На выплату выигрышей будет потрачено 1500 рублей.

Таким образом, доход от лотереи составит 500 рублей.

Разумеется, может случиться так, что на один купленный нами билет мы получим большой выигрыш. Но если бы некто решил купить все билеты, то он достоверно потерял бы 500 рублей — по 5 на каждый из 100 билетов.

 Так устроены все лотереи: математическое ожидание выигрыша на один билет меньше цены этого билета.

Это условие является непременным, и оно обеспечивает рентабельность лотереи и доход ее устроителям. Человек, который решил сыграть в лотерею, должен понимать это и сознательно рисковать своими деньгами.

**Домашнее задание:**

1. Написать краткий конспект;
2. Решить задачи:

**Задача1**. В небольшой фирме 10 сотрудников: 7 рабочих, мастер, бухгалтер, директор. Зарплата у рабочих: 2000, у мастера 4000, у бухгалтера 16000, у директора 40000. Найдите, чему будет равна средняя зарплата на этом предприятии?

**Задача 2**. В таблице приведена информация о длине некоторых рек, протекающих по территории Ивановской области, впадающих в реку Волга

|  |  |
| --- | --- |
| Река | Длина (км) |
| Елнать | 54 |
| Кинешемка | 34 |
| Казоха | 9 |
| Мера | 152 |
| Шача | 58 |
| Солоница | 132 |
| Сунжа | 45  |

а) Составьте ранжированный ряд;

б)Найдите среднюю длину рек (среднее арифметическое);
в)Найдите длину рек в среднем (медиану данных);
г) Найти размах длины рек.

**Задача 3.**В женском обувном магазине провели статистические исследования и составили соответствующую таблицу по цене обуви и количества продаж:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Цена (руб.)** | 500 | 1200 | 1500 | 1800 | 2000 | 2500 |
| **Количество** | 8 | 9 | 14 | 15 | 3 | 1 |

Первый и второй этап статистического исследования уже пройдены: данные собраны и систематизированы. Осталось произвести анализ данных. Для данных показателей надо найти статистические характеристики.