

Бронникова Лариса Михайловна, канд. педагогических наук, доцент

Теория вероятностей и статистика.

Описательная статистика.

Раздел «Вероятность и статистика» – обязательный компонент школьного образования. Этот материал необходим для формирования у учащихся умения воспринимать и критически анализировать информацию, представленную в различных формах, понимать вероятностный характер многих реальных зависимостей, производить простейшие вероятностные расчеты.

Представление данных и описательная статистика в ФГОС ООО распределены на 7-9 классы следующим образом:

7 класс – представление данных в виде таблиц, диаграмм, графиков, извлечение информации, использование и интерпретация данных, описательная статистика, среднее значение;

8 класс – представление данных в виде таблиц, диаграмм, графиков, множество, элемент множества, подмножество, математическое описание случайных величин.

9 класс – представление данных в виде таблиц, диаграмм, графиков, чтение и построение таблиц и графиков по реальным данным.

В школьном курсе изучение основных понятий статистики, очевидно, должно проходить на элементарном, интуитивном уровне. Поэтому авторы ряда основных школьных учебных пособий предлагают начинать знакомство с разделами теории вероятностей и статистики именно со статистики. Такой подход исходит из нескольких методических соображений.

Во-первых, большая часть материала описательной статистики (средние значения, медиана, размах) доступна для большинства учащихся 7-9-х классов.

Во-вторых, при обсуждении реальных статистических данных хорошо иллюстрируется случайная изменчивость окружающего нас мира. Тем самым

готовится наглядная концептуальная база для понятий «случайный эксперимент» и «вероятность исхода» этого эксперимента.

В-третьих, на жизненном статистическом материале школьники видят, как формализуются и описываются данные.

В-четвертых, у учителей появляется возможность на неформальных примерах повторить и закрепить ряд тем школьного курса математики (доли и проценты, мера угла, навыки арифметических действий с числами разных знаков).

Изучение раздела «Описательная статистика» начинается со знакомства с таблицами и диаграммами. Аргументируется, что таблицы нужны, чтобы упорядочивать большие массивы данных. Гораздо легче искать информацию в таблице, чем в обычном тексте, потому что в таблице каждое значение находится в своей ячейке, а однородные сведения сгруппированы в одной графе. Если данные подходящим образом помещены в таблицу, то их удобно сравнивать.

К графическим способам представления статистических данных в школьном курсе, прежде всего, относятся диаграммы – столбиковая, круговая и диаграмма рассеивания. По внешнему виду столбиковую диаграмму часто путают с другим графиком – гистограммой, отражающим совершенно другие характеристики набора чисел.

Термин «гистограмма» имеет четко очерченный смысл. Гистограмма является частным случаем столбиковой диаграммы, строится специальным образом для сгруппированных исходных наборов чисел. При построении гистограммы ось абсцисс делится на интервалы, чаще всего одинаковые. С каждым интервалом группировки данных мы связываем частоту попадания чисел из заданного набора в указанный интервал. Эта частота и откладывается на графике гистограммы по оси ординат. В отличие от таблиц, диаграммы не передают информацию совершенно точно, зато они наглядны и позволяют сравнивать величины быстро – на глаз.

Иногда приходится сопоставлять две разные величины, интересуясь, есть ли связь между ними и что это за связь. В этом помогает специальный инструмент – диаграмма рассеивания. На плоскости диаграммы отмечают точки с двумя координатами; абсцисса соответствует одной изучаемой величине, ордината – другой. Рассматривая полученные точки, можно делать разумные предположения.

Если чем больше одна величина, тем больше другая, то такая связь называется положительной. Если же с увеличением одной величины другая уменьшается, связь называется отрицательной. Кроме того, связи бывают линейными и нелинейными. Нелинейные связи изучать сложнее, чем линейные. Чтобы выразить линейную связь численно, надо провести прямую линию сквозь облако рассеивания и установить коэффициенты этой прямой. В математике есть специальные методы построения такой прямой, но в школьном курсе делают это на глаз, стараясь, чтобы прямая «протыкала» облако ровно посередине и чтобы точки прилегли к ней как можно плотнее.

Информационная культура давно ввела в оборот понятие «генеральной совокупности». Оно употребляется в средствах массовой информации. Однако нет необходимости обсуждать это понятие в школьном курсе во всей полноте. Этот термин, если он употребляется учителем или встречается в учебных пособиях, должен использоваться для обозначения только **конечного множества всех** обсуждаемых объектов.

На практике обо всей генеральной совокупности часто судят лишь по какой-то ее части. В курсе статистики основной школы мы называем «выборкой» часть объектов, выбранных из конечной генеральной совокупности, не оговаривая каких-либо требований к самому способу отбора. На наш взгляд, не следует упоминать выборку без указания генеральной совокупности, из которой она получена. У выбранных объектов может быть измерена какая-то их числовая характеристика. Так получается «набор чисел», который можно считать выборкой из рассматриваемой генеральной совокупности.

Набор чисел не нужно никак определять. Как правило, под набором понимается неупорядоченное множество чисел, при этом числа в наборе могут повторяться. Упорядоченные наборы (например, временные ряды) являются гораздо более сложными объектами и требуют специальных методов работы с ними. Представляется не вполне удачным называть набор чисел словом «данные» и использовать это слово как термин. С одной стороны, в обыденной речи слово «данные» имеет размытый и неопределенный смысл. С другой стороны, выражение «анализ данных» прочно закрепилось как собирательное название ряда разделов математической статистики.

Затем школьники знакомятся с числовыми характеристиками набора: знакомятся с тем, как можно числами описывать, насколько данные рассеяны относительно своего среднего, учатся вычислять дисперсию и стандартное отклонение; знакомятся с основными свойствами отклонений, дисперсии и стандартного отклонения, знакомятся с диаграммами рассеивания.

При знакомстве с описательными статистики числовые данные сначала рекомендуется подходящим способом представить, а затем – описать. Однородные данные удобно описывать одним числом. Для этого в разных случаях используются среднее арифметическое, медиана, наибольшее или наименьшее значение и другие характеристики. Выбор подходящей характеристики делают, исходя из цели исследования и природы данных.

Под размахом понимают меру рассеивания данных. Он показывает, как далеко отстоят друг от друга наибольшее и наименьшее значения. Размах часто используется в грубых бытовых измерениях.

Среднее арифметическое набора чисел вычисляют как отношение суммы чисел к их количеству. Оно имеет ясный физический смысл – «точка равновесия» данных на числовой прямой. Его разумно применять в качестве меры центра для суммируемых величин (объем производства, доходы, расходы, численность населения и т.п.). Однако среднее арифметическое не

подходит для прогнозов в случае, когда величина сильно меняется во времени (имеет тенденцию к росту или падению).

Число m называется медианой числового набора, если в этом наборе хотя бы половина чисел не больше числа m и хотя бы половина чисел не меньше числа m . Медиана набора числовых данных иногда определяется неоднозначно. Но на практике в таких случаях в качестве медианы используют среднее арифметическое двух срединных чисел.

Под модой понимается самое часто встречаемое значение в наборе. Из этого определения ясно, что мода может не быть однозначно определена. Поэтому в определении моды требуются уточнения. Для наборов, где каждое значение встречается только один раз или одинаковое число раз (скажем, два), говорят, что мода отсутствует. Если несколько значений в наборе (но не все) встречаются с одинаковой наибольшей частотой, то говорят, что мода принимает несколько значений.

Разные источники предлагают слово «частота» для обозначения то абсолютной, то относительной частоты. Это различие в толковании термина порождает серьезную путаницу.

В теории вероятностей частота нужна для частотного определения вероятности события, и здесь речь идет, конечно, о частоте относительной. Абсолютная частота редко употребляется в современной литературе, и поэтому использование термина «частота» в этом смысле уже можно считать архаизмом.

В соответствии с практикой последних десятилетий, мы предлагаем использовать в школьном курсе слово «частота» только в смысле относительной частоты.

Итак, частота – это отношение числа наступлений события к общему числу наблюдений или, другими словами, доля события в наборе наблюдений.

Термин «абсолютная частота» лучше не употреблять совсем, а говорить о числе наступлений интересующего нас события.

Рекомендуется говорить «среднее набора», «медиана набора» или «дисперсия набора» и т.д., не использовать эти слова в отрыве от набора, так как в дальнейшем эти же термины будут использоваться для описания случайных величин. Будет говориться о среднем, медиане и дисперсии случайной величины.

Сходство терминов не случайно. Перечисленные характеристики набора чисел дают нам приближенное представление об аналогичных характеристиках случайной величины. Эта взаимосвязь является очень важной в математической статистике.

Вводятся определения среднего геометрического и среднего гармонического набора, изучаются отклонения и их основные свойства, распределения случайной величины и его числовые характеристики – математическое ожидание и дисперсия. Под дисперсией понимают средний квадрат отклонений. Дисперсия – очень важный инструмент для описания числовых наборов. В Excel существует специальная функция для вычисления дисперсии.

Стандартным отклонением набора чисел называется квадратный корень из дисперсии этого набора. Для стандартного отклонения тоже есть специальная функция в Excel.

Отметим, что аппарат математической статистики является изумительным по мощности и гибкости инструментом для отсеивания закономерностей от случайностей. Описательная статистика предназначена для представления данных в удобном виде и описания информации в терминах математической статистики и теории вероятностей.