МАТЕМАТИЧЕСКАЯ ВЕРТИКАЛЬ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И СТАТИСТИКА. 7 КЛАСС (2 ч./нед.)

Раздел «Описательная статистика»

Урок 15. Числовые наборы.
Меры центральной тенденции (центральные меры, средние)

Материалы представляют собой примерный сценарий урока по теме «Числовые наборы. Меры центральной тенденции». Учитель может на свое усмотрение использовать сценарий целиком или частично, используя фрагменты наряду с собственными разработками и материалами учебника[1]*. Авторы будут благодарны за замечания и предложения по структуре и содержанию сценариев.*

Цель урока – сформировать у учащихся общее представление об описании числовых данных с помощью одной числовой характеристики - центральной меры (среднего). Учащиеся должны получить представление о том, что центральных мер много и способ описания данных в каждой задаче выбирается с учетом трех основных факторов:

- природа данных;
- цель исследования;
- сложившиеся традиции.

Что такое описательная статистика. Когда данные собраны и представлены в удобном виде, например, с помощью таблицы, возникает вопрос: где и как группируются эти данные – велики или малы значения и каков разброс (рассеивание) данных. Чтобы понять это, данные нужно *описать подходящим образом.* Для этого нужны *описательные показатели или параметры*: средние значения, наибольшее, наименьшее и другие. Поэтому раздел статистики, занимающийся описанием данных, называется *описательной статистикой*.

Наборы числовых данных – генеральная совокупность, выборка, ряд. Числовые данные могут быть устроены сложно. Они могут иметь разную природу и требовать разных способов описания. Например, бывают *генеральные совокупности данных*, бывают выборки, часто встречаются *ряды*.

[1] Математика 7-9 класс. Теория вероятностей и статистика / Ю. Н. Тюрин, А. А. Макаров, И. Р. Высоцкий, И. В. Ященко. – 3-е изд., стереотипное. – М.: МЦНМО: ОАО «Московские учебники», 2011. – 256 с.: ил.

Пример 1. Что такое *генеральная совокупность*? Предположим, нам нужно собрать сведения о росте жителей Российской Федерации. Сразу возникают проблемы: у взрослых и у детей разный рост. Женщины и мужчины тоже отличаются по росту. Уточним задачу: мы хотим изучить рост только взрослых женщин. В этой задаче все взрослые жители России женского пола образуют генеральную совокупность. Таким образом, генеральная совокупность – это набор всех объектов, которые нужно исследовать.

Что такое *выборка?* Измерить рост всех женщин в России невозможно. Их миллионы. Собрать такую статистику – чрезвычайно трудно и дорого. Но можно выбрать из совокупности часть – *выборку*. В нашем случае, можно сделать выборку объемом 1000, то есть случайным образом выбрать 1000 женщин и измерить их рост. Это вполне посильная задача. Чтобы по выборке сделать достоверный или хотя бы правдоподобный вывод о всей совокупности, используются специальные *выборочные методы*. Например, если в нашей выборке средний рост равен 165 см, можно считать, что средний рост всех женщин во всей генеральной совокупности близок к этому значению[2].

Обсудите со школьниками, кому и зачем важно иметь сведения о росте взрослых женщин? Можно ли поменять местами порядок чисел в выборке? Изменится ли от этого среднее арифметическое, наибольшее или наименьшее значение в выборке?

Пример 2. Что такое *ряд*? Предположим, что мы следим за некоторой изменчивой величиной. Например, за ценой на нефть. Цена нефти выражается в долларах за баррель[3]. Каждый день фиксируется средняя цена за 1 баррель нефти марки Brent[4]. В таблице показаны цены за 10 дней. Если данные каким-то образом упорядочены (по времени, по величине и т.п.), они образуют *ряд*.

Дата	Ср.цена	Изм.
29.10.2018	76,92	-1,08%
30.10.2018	76,32	-0,78%
31.10.2018	76,23	-0,12%
01.11.2018	72,75	-4,57%
02.11.2018	72,61	-0,19%
06.11.2018	72,1	-0,70%
07.11.2018	72,05	-0,07%
08.11.2018	70,69	-1,89%
09.11.2018	69,88	-1,15%
12.11.2018	70,94	1,52%

Кому важно знать, как менялась цена на нефть во времени? Как можно использовать эту информацию? Можно ли в этой таблице поменять местами цены за разные дни?

Желательный результат обсуждения. В примере 1 мы имеем дело с выборкой. Важно знать не только среднее, но и сколько процентов женщин не выше 150 см, 155 см и так далее. В первую очередь, эта информация важна для модельеров, производителей одежды и мебели. А вот порядок измерений здесь

[2] Насколько среднее в выборке близко к среднему всей генеральной совокупности – важный вопрос теории вероятностей.

[3] Баррель – от английского barrel (бочка) – мера объема, равная примерно 159 л.

[4] Brent – нефтяная смесь из нескольких месторождений Северного моря. Цена на нефть марки Brent задает ценовой стандарт для стран Европы. Существуют другие стандарты. В России употребительным стандартом является цена на нефть марки Urals.

не важен. Ни среднее арифметическое, ни наибольшее и наименьшее значения не поменяются от перестановки измерений.

Информация о ценах на нефть важна для нефтяников, экономистов и для правительства. Например, бюджет России в значительной степени формируется из доходов от продажи нефтепродуктов. Информация об изменении цен во времени – динамике – позволяет строить прогнозы, сравнения, принимать регулирующие решения. В этом примере важна не только сама цена, но и как она менялась день ото дня. Ряд – упорядоченный набор данных. Изменение порядка данных приведет к потере важной информации и даст нам другой ряд с другими свойствами.

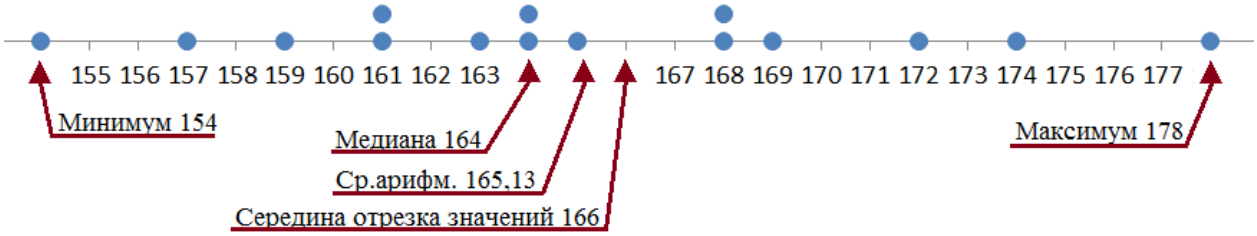
Вывод. Выборки и ряды – специальные виды данных. Они требуют особых способов описания и обработки. Поэтому мы не будем специально изучать выборки и ряды, но все равно при изучении данных разной природы нам придется обсуждать, можно ли применить те или иные способы описания.

Центральные меры (средние, меры центральной тенденции)

Рассмотрим какой-нибудь числовой набор данных. Пусть это будет рост 15 случайно выбранных взрослых женщин в сантиметрах:

154, 157, 159, 161, 161, 163, 164, 164. 165, 168, 168, 169, 172, 174, 178.

Для наглядности нанесем эти значения на координатную прямую.



Мы видим, что все значения расположены на отрезке от 154 до 178 см. Середина отрезка значений - 166 см. Легко найти среднее арифметическое: приближенно 165,13 см. Позже мы подробно изучим медиану, а сейчас просто скажем, что медиана этих данных равна 164 см. Все эти значения (и вообще любое значение от 154 до 178 см) можно рассматривать как *центральную меру* данного набора. Вместо слов «центральная мера» часто используют просто слово «*среднее*» (не обязательно при этом имеется в виду среднее арифметическое)[5].

В одних случаях удобны одни средние, в других случаях – другие. Обсудите со школьниками, в каких случаях важно знать наибольшее или наименьшее значение. Чем удобна середина отрезка? Чем хорошо среднее арифметическое?

[5] В таком расширенном смысле даже наибольшее и наименьшее значение являются средними. Хотя, конечно, говоря о среднем, мы обычно не подразумеваем крайние значения.

Желательные результаты обсуждения. Если нас интересует самый высокий человек, то нужно ориентироваться на максимум (например, при проектировании дверей в домах). Если мы проектируем офисную мебель, то нужно опираться и на максимум, и на минимум (стул должен регулироваться так, чтобы даже самая низкорослая сотрудница доставала ногами до пола). Середина отрезка очень удобна, поскольку очень легко и быстро вычисляется. Но она зависит только от наибольшего и от наименьшего значений, а потому является ненадежным средним – слишком сильно подвержена случайной изменчивости. Среднее арифметическое надежнее – оно зависит от всех значений в наборе. Но оно также имеет недостатки: например, не для всех данных оно имеет смысл.

Пример 3. Сергей сдавал зачет по прыжкам в длину с места. В трёх попытках у него получились следующие результаты:

165 см, 140 см, 178 см.

Какую «центральную меру» учитель физкультуры впишет в журнал против фамилии Сергея? Имеет ли смысл здесь вычислять среднее арифметическое?

Пример 4. Маша сдавала зачет по бегу на 60 м. У нее в трех попытках были следующие результаты: 11,3 с, 9,9 с и 10,1 с.

Какую «центральную меру» учитель физкультуры впишет в журнал против фамилии Маши? Имеет ли тут смысл среднее арифметическое в этом примере?

Желательный результат обсуждения. Несколько попыток нужно, чтобы выбрать лучшую. В журнал вписывается лучший результат, а не средний. В примере 3 это максимум 178 см, а в примере 4 это минимум 9,9 с. Выбор описательной характеристики зависит от природы данных (в одном случае это время, а в другом – дальность прыжка) и от цели исследования (в обоих примерах – это поиск лучшего результата).

Пример 5. У Васи по математике за четверть накопилось 10 отметок:

2, 3, 2, 4, 4, 3, 4, 2, 5, 3.

Учитель ставит четвертную отметку по обычному правилу: находит среднее арифметическое и округляет до целой отметки. Имеет ли смысл, на ваш взгляд, такой способ выведения итоговой отметки с точки зрения математики?

Желательный результат обсуждения. Школьные отметки, к которым мы привыкли – это цифры, которыми в 1947 году заменили словесные отметки «плохо», «посредственно», «хорошо» и «отлично». Можно ли найти среднее арифметическое словесных отметок? Конечно, нет: непонятно, как складывать слова. Цифровые отметки складывать можно, но смысла от этого не приба-

вится: двойка плюс пятерка не даст отметку семь. Поэтому же в школах среднюю отметку считают, пользуясь средним арифметическим? Здесь есть три причины, и неясно, какая главная.

1. Среднее арифметическое в равной степени учитывает все отметки;
2. Среднее арифметическое очень просто вычислять;
3. Так повелось издавна, и все к этому привыкли.

Поэтому с отсутствием «математического смысла» средней арифметической отметки все смирились.

Выводы. Числовые данные сначала нужно подходящим способом представить, а затем – описать. Однородные данные удобно описывать одним числом – центральной мерой или мерой центральной тенденции, или, попросту говоря, каким-либо средним. Таких характеристик множество. Например, среднее арифметическое, минимум, максимум, середина отрезка значений. Выбор среднего делают, учитывая три обстоятельства.

1. Природа данных (в одном случае естественно вычислять среднее арифметическое, в другом – максимум и т.п.)
2. Цель исследования (вспомните примеры по проектирование дверей, мебели и спортивные рекорды).
3. Сложившаяся традиция (может быть, полезно найти что-то другое, но все привыкли делать по-старому, и ничего не хочется менять).

Рекомендуемое домашнее задание: см. Приложение.

Дополнительные материалы для подготовки к уроку. Учитель на свое усмотрение может свободно использовать таблицы[6], размещенные на сайте «Вероятность в школе» http://ptlab.mccme.ru/node/350.
При проведении урока в классе, оборудованном компьютерами на столах учащихся, учитель может включить в урок вычисления в электронных таблицах непосредственно на рабочих столах учащихся.

[6] Все данные взяты из открытых источников и представлены в формате .xls или .xlsx.Таблицы можно использовать как целиком, так и делая из них необходимые выдержки.

Приложение

1. Перед проектированием железнодорожного моста через реку изыскатели собирают данные о паводковом уровне этой реки за много лет. Для чего нужны эти данные? Какая мера собранных данных важна в этом случае?

Предполагаемый ответ. Нужно знать наибольший уровень. Мост должен быть выше максимального уровня подъема воды.

2. Комментатор футбольного матча часто сообщает слушателям средний возраст игроков команды (среднее арифметическое). Подумайте, является ли среднее арифметическое возрастов игроков подходящим описанием возраста «типичного игрока команды». Обоснуйте свое мнение.

Предполагаемый ответ. Среднее арифметическое здесь – дань традиции. Ведь возраст игрока - не суммируемая величина.

3. Во многих случаях (например, при рассмотрении заявки на кредит или заявления на визу) работающего гражданина спрашивают о среднем доходе за последние три месяца. Как вы думаете, почему не имеет смысла спрашивать о среднем доходе за более длительный период? Обоснуйте свое мнение.

Предполагаемый ответ. Цель такого опроса – понять, насколько гражданин платежеспособен в настоящее время. Включение в опрос давних зарплат может исказить картину. Например, раньше человек зарабатывал очень много, но потерял работу. Среднее большое, но не отражает нынешнее состояние дел. Или наоборот, полгода назад гражданин устроился на высокооплачиваемую работу, и сейчас не испытывает трудностей, но старые данные покажут низкое среднее, не соответствующее текущему благосостоянию.

4. Гражданин П. собрал все квитанции об оплате ЖКУ (квартплата, электричество, вода) за последние три года и нашел среднее арифметическое всех выплат. Может ли П. ориентироваться на полученное значение в будущем? Обоснуйте свое мнение.

Предполагаемый ответ. К сожалению, тарифы на ЖКУ растут. Среднее за три предыдущих года никак не описывает будущие тарифы: оно может оказаться ниже, чем наименьший платеж в наступающем году. Гражданин П. выбрал неудачный способ описания временного ряда.