О пользе статистики

Н.Н. Осипенко

Автор: Осипенко Наталья Николаевна, учитель математики и информатики средней школы № 6 г. Партизанска Приморского края.

Предмет: Алгебра.

Класс: 7.

Тема: Статистические характеристики.

Профиль: Общеобразовательный.

Уровень: Общий.

Текст задачи. На некотором предприятии проводился статистический анализ заработной платы работников. Были получены следующие результаты: среднее арифметическое 17,6, медиана 13, мода 12, размах 67. Что обозначают эти характеристики? Какие из них несут полезную информацию о заработной плате? В каких ситуациях целесообразно использовать в качестве обобщающего показателя среднее арифметическое, медиану, моду, размах?

а) Выделите ключевые слова для информационного поиска.

б) Найдите и соберите необходимую информацию.

в) Обсудите и проанализируйте собранную информацию.

г) Сделайте выводы.

д) Сравните ваши выводы с культурным образцом.

Возможные информационные источники

Книги:

Тюрин Ю.Н. и др. Теория вероятности и статистика. 2-е изд., перераб. М.: МЦНМО: ОАО «Московские учебники», 2008.

Шевелева Н.В. Математика (алгебра, элементы статистики и теории вероятностей). 9 класс / Н.В. Шевелева, Т.А. Корешкова, В.В. Мирошин. М.: Национальное образование, 2011. (Краткий курс).

Web-сайты:

http://teorver.mccme.ru/tmvy/book/3/

http://www.docme.ru/doc/4071/%D1%81%D1%82%D0%B0% D1%82%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0

http://pubhealth.spb.ru/EpidD/epidD3.htm

http://art.ioso.ru/seminar/2009/projects11/rezim/stat4.html

Культурный образец

Бродский Я.С. Статистика. Вероятность. Комбинаторика. М.: Оникс; Мир и Образование, 2008 (Школьный курс математики).

В результате исследований, связанных с массовыми явлениями, получают много числовых данных. Возникает проблема — найти такие характеристики, которые довольно полно характеризовали бы полученный числовой материал. Характеристики, которые базируются на данных массовых наблюдений, называют обобщающими показателями. Эти показатели характеризуют значения признака, его вариацию. Важнейшие среди обобщающих показателей — средние величины, т.е. такие значения признака, вокруг которых группируются отдельные наблюдаемые значения элементов.

В зависимости от характера задачи пользуются тем или иным видом средней величины. К ним принадлежат среднее арифметическое, мода, медиана, степенные средние (среднее гармоническое, среднее геометрическое и т.п.).

Изучая и используя обобщающие показатели, следует иметь в виду, что они только тогда объективно будут соответствовать своему назначению, если применяются к однородным совокупностям. В противном случае можно получить неправильные выводы.

С понятием среднего арифметического вы уже знакомы с младших классов. Напомним его. Пусть имеется n объектов, для которых измерена некоторая характеристика и получены значения $x_1, x_2, \ldots, x_n$. Среднее арифметическое этих n значений обозначают через $\bar{x}$ и определяют как

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$, или

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Символом $\sum_{i=1}^{n} a_i$ с переменным индексом *i* (читается так: «сумма по *i* от 1 до *n*») обозначают следующую сумму:

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \ldots + a_n.$$

Сущность среднего арифметического состоит в следующем. Если каждое наблюдение заменить средним, то общая сумма не изменится. Это среднее можно интерпретировать ещё и так: если все наблюдения будут равны между собой, а сумма наблюдений останется неизменной, то каждое наблюдение будет равно среднему. Поскольку среднее сохраняет неизменной сумму при равномерном распределении значений, то оно наиболее полезно в качестве обобщающего показателя при отсутствии резко выделяющихся наблюдений, или, как их называют, выбросов, т.е. когда набор данных представляет собой более-менее однородную группу.

Пример 1. Рассмотрим среднюю месячную зарплату работников некоторого предприятия. Пусть, например, в фирме работает 20 человек, зарплата 19 из них составляет 10 000 руб., а зарплата 20-го, руководителя, 1 000 000 руб. Тогда средняя зарплата одного работника на этой фирме будет равна

$$\frac{19 \cdot 10000 + 1000000}{20} = 59500.$$

Хотя среднее и сохранило общую сумму заработной платы, но оно является в данном случае плохим обобщающим показателем: оно плохо характеризует зарплату одного работника на этой фирме. Причина этого кроется в том, что набор данных содержит выброс: 1000000 руб. Среднее оказалось слишком большим для большинства работников и слишком малым для высокооплачиваемого руководителя.

Пример 2. Два стрелка сделали по 100 выстрелов. Первый выбил 8 очков 40 раз, 9 очков 10 раз и 10 очков 50 раз.

Второй выбил 8, 9 и 10 очков соответственно 10, 60 и 30 раз. Какой из стрелков стреляет лучше?

Вычислим средние арифметические x и y числа очков, которые выбил при 100 выстрелах каждый из двух стрелков.

$$\bar{x} = \frac{8 \cdot 40 + 9 \cdot 10 + 10 \cdot 50}{100} = 9{,}1;$$

$$\bar{y} = \frac{8 \cdot 10 + 9 \cdot 60 + 10 \cdot 30}{100} = 9{,}2.$$

Среднее число очков, которое выбивает из 100 выстрелов второй стрелок, несколько выше, чем тот же показатель у первого стрелка. Естественно признать второго стрелка лучшим.

Кроме среднего арифметического мерой центральной тенденции может служить:

1) медиана, или срединная точка, которую можно вычислять как для порядковых, так и для количественных данных;

2) мода — наиболее часто встречающаяся категория, которую можно вычислять для номинальных данных, упорядоченных категорий и количественных данных.

Среднее арифметическое плохо характеризует состояние с заработной платой на предприятии. Средний арифметический заработок оказался слишком высоким для работников предприятия. Более правильную картину даст то значение, которое делит данные на две равные части. Таким значением является медиана.

Если все элементы совокупности размещены в порядке возрастания или убывания числовых значений признака, то медиана — это такое значение признака, которое делит всю совокупность пополам.

Будем обозначать медиану символом Ме.

При нахождении медианы дискретного вариационного ряда следует различать два случая:

1) объём совокупности нечётный;

2) объём совокупности чётный.

Если объём совокупности нечётный и равен $2n+1$, и варианты размещены в порядке возрастания их значений:

$$\underbrace{x_1, x_2, \ldots, x_n}_{n \text{ значений}}, x_{n+1}, \underbrace{x_{n+2}, \ldots, x_{2n+1}}_{n \text{ значений}},$$

то Ме = x_{n+1}.

Если же количество элементов чётное и равно 2n, то нет варианты, которая бы делила совокупность на две равные по объёму части:

$$\underbrace{x_1, x_2, \ldots, x_n}_{n \text{ значений}}, \underbrace{x_{n+1}, \ldots, x_{2n}}_{n \text{ значений}}.$$

Поэтому в качестве медианы условно берется полусумма вариант, находящихся в середине вариационного ряда:

$$Me = \frac{x_n + x_{n+1}}{2}.$$

Пример 3. Вычислить медиану по данным таблицы 1, в которой приведена информация об успеваемости по математике 100 учащихся 7-х классов (успеваемость оценивается по 12-балльной шкале).

Поскольку совокупность содержит 100 элементов (учащихся), упорядоченных по значению признака, и количество элементов чётно, то надо найти полусумму числовых значений 50-го и 51-го элементов. Складывая последовательно частоты, находим накопленные частоты (табл. 2).

Устанавливаем первую накопленную частоту; она больше полови-

Таблица 1

Количество баллов	1	2	3	4	5	6	7	8	9	10	11	12
Число учащихся	3	4	4	9	11	12	18	14	9	8	6	2

Таблица 2

Количество баллов	1	2	3	4	5	6	7	8	9	10	11	12
Накопленные частоты	3	7	11	20	31	43	61	75	84	92	98	100

Таблица 3

Количество баллов	1	2	3	4	5	6	7	8	9	10	11	12
Число учащихся	1	1	2	3	4	4	6	5	3	3	2	1

ны общего количества элементов. В данном примере это 61. Итак, и 50-му, и 51-му элементам отвечает значение 7, поэтому и Ме=7.

Полученное значение медианы означает, что примерно половина семиклассников по математике учатся на 7 и меньше баллов, а половина — на 7 и больше баллов.

Обращаем внимание на ошибку, часто встречающуюся при вычислении медианы. Иногда не учитывают ни частоты вариант, ни общего количества элементов и в качестве медианы берут полусумму средних вариант. В примере 1 это полусумма 6-й и 7-й вариант, т. е. 6,5, что неверно.

Медиана обладает важными свойствами, которые в некоторых случаях дают ей преимущество перед другими средними величинами. Например, если при упорядоченном размещении некоторого признака «крайние» значения сомнительные и к тому же резко отличаются от основной массы данных, то в качестве меры центральной тенденции целесообразно использовать медиану, поскольку на её величину эти «крайние» значения никакого влияния не оказывают, и в то же время они могут существенным образом повлиять на значение среднего арифметического.

Выше рассматривался пример с зарплатой работников некоторой фирмы, в которой работает 20 человек, зарплата 19 из них составляет 10 000 руб., а зарплата 20-го, руководителя, 1 000 000 руб. Как мы видели, средняя зарплата одного работника на этой фирме составляет 59 500 руб. Среднее арифметическое явилось в данном случае плохой мерой центральной тенденции. Медиана данной совокупности равна 10 000 руб. Она лучше характеризует совокупность, состоящую из размеров зарплат работников фирмы.

Мода — это такое значение признака, которое встречается наиболее часто. В случае дискретных рядов вычислить моду нетрудно. Достаточно найти варианту, которая имеет наибольшую частоту или относительную частоту, это и будет мода. Будем обозначать моду символом Мо.

Пример 4. В таблице 3 приведены итоговые оценки учащихся некоторого класса по математике. Найти моду данного распределения.

Из всех оценок чаще всего встречается 7 баллов: шесть раз. Поэтому Мо = 7. Этот результат имеет вполне определённый смысл — больше всего учащихся класса имеют по математике 7 баллов.

Если все значения в вариационном ряде встречаются одинаково часто, то считают, что этот ряд не имеет моды.

Если два соседних значения вариационного ряда имеют одинаковую частоту и она больше частоты любого другого значения, то считают, что мода равняется среднему арифметическому этих значений.

Если два несоседних значения вариационного ряда имеют одинаковую частоту и она больше частоты любого другого значения, то считают, что вариационный ряд имеет две моды, а соответствующее распределение называют бимодальным.

Средние величины являются важными характеристиками статистических совокупностей. Они говорят нам о концентрациях совокупности значений на числовой шкале. Каждая мера центральной тенденции даёт такое значение, которое «представляет» в определённом смысле все значения совокупности. В этом случае пренебрегают различиями, существующими между отдельными значениями. Для измерения разброса и вариации значений внутри совокупности нужны другие показатели.

Пример 5. Пусть средний рост учащихся 8 «А» класса равен 1,65 м, а 8 «Б» — 1,7 м. По этим данным нельзя определить, в каком из этих двух классов учится самый высокий ученик.

Размах измеряет на числовой шкале расстояние, в пределах которого изменяются значения совокупности. Различают два вида размаха. Остановимся на одном.

Исключающий размах — это разность между максимальным и минимальным значениями в совокупности.

Пример 6. Вычислить исключающий размах по приведённым данным о диаметрах валиков (в мм):

14,51	14,42	14,56	14,47	14,46
14,35	14,48	14,53	14,21	14,31
14,35	14,68	14,56	14,28	14,36
14,21	14,52	14,23	14,41	14,46
14,69	14,54	14,36	14,15	14,37
14,51	14,25	14,55	14,51	14,36
14,62	14,55	14,38	14,33	14,40
14,52	14,48	14,51	14,55	14,39
14,54	14,58	14,48	14,37	14,38
14,51	14,36	14,15	14,24	14,32

Наименьшее значение диаметра равно 14,15 мм, наибольшее — 14,69 мм. Исключающий размах равен 14,69 — 14,15 = 0,54.

Преимуществом этого показателя является очевидная простота его вычисления. Но часто он даёт лишь приближённую характеристику вариации. Это особенно заметно в случае достаточно больших по объёму совокупностей, когда подавляющее большинство элементов имеют значения признака, компактно сгруппированные вокруг некоторой средней величины, и только некоторые из них в силу случайных обстоятельств принимают значения (наибольшее и наименьшее), существенно отличающиеся от основной массы. При этом размах вариации будет значительным, а вариация по существу малой. Дело в том, что при вычислении размаха не учитывается каждое отдельное значение.

Методический комментарий

Задача решается при изучении темы «Статистические характеристики». Цель решения задачи — выяснить, что представляют собой основные статистические характеристики, что они характеризуют, в каких ситуациях их целесообразно использовать.