Тема № 30 «Табличное и графическое представление данных».

Когда сведений очень много, их нужно упорядочивать. ***Таблица*** – самый простой способ упорядочить данные. С некоторыми таблицами вы уже имели дело. Это таблицы сложения и умножения чисел, таблицы спряжения глаголов. Таблицами являются: расписание уроков, страницы школьного дневника, оглавление учебника. Таблицы облегчают поиск необходимых сведений, не заставляя изучать всю имеющуюся информацию.

Однако таблицы не дают наглядного представления о соотношении величин. Для этого служат различные диаграммы: столбиковые, круговые, рассеивания и др. Пословица гласит: «Лучше один раз увидеть». ***Диаграммы*** используются для наглядного, запоминающегося изображения и сопоставления данных.

Таблицы и диаграммы удобно применять для сравнения шансов случайных событий, используя статистические данные (числовые данные, полученные в результате различных наблюдений, опросов, экспериментов.)

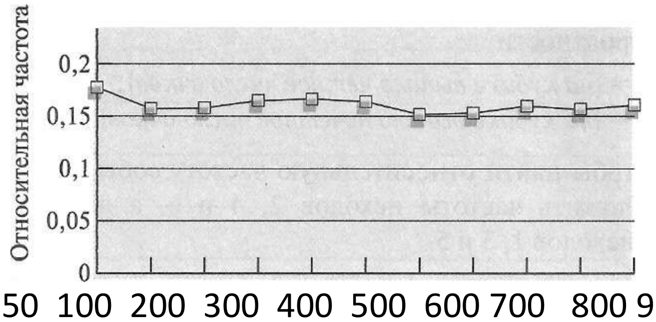
На основе собранных статистических данных вычисляют частоту случайного события и выясняют, как она связана с вероятностью.

Чтобы выяснить, насколько вероятно то или иное случайное событие, связанное со ***случайным экспериментом*** (эксперимент, условиями которого являются непредсказуемость и возможность многократного повторения), нужно подсчитать, как часто оно происходит. Для этого используют две важные величины – абсолютную и относительную частоту.

Абсолютная частота показывает, сколько раз в серии экспериментов наблюдалось данное событие. Это всегда целое число.

Относительная частота (частота) показывает, какая доля экспериментов завершилась наступлением данного события. Она равна отношению n - числа опытов, в которых интересующее нас событие произошло, к N - общему числу проведенных опытов. Это дробное число, меняющееся от 0 до 1.

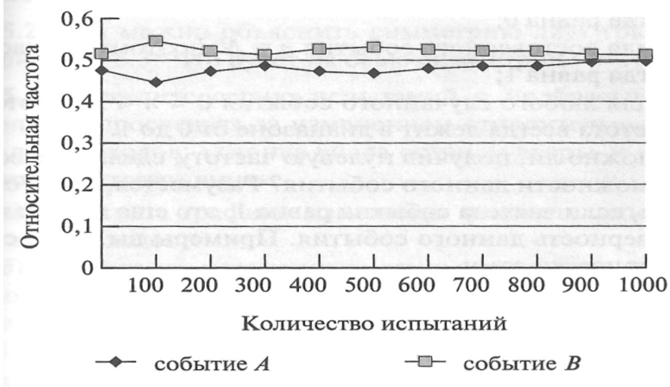
Опытным путем установлено, что проводя эксперименты огромное количество раз (больше 1000), например, такие эксперименты, как бросание игральной кости, бросание монеты, раздача игральных карт, розыгрыш лотереи и др., частоты становятся устойчивыми. Например, изобразим график зависимости частоты от числа опытов при бросании игральной кости, показывающий как часто выпадала «единица»:



50 100 200 300 400 500 600 700 800 900 1000

Количество экспериментов

Другой пример. Событие А – «на кубике выпало четное число очков», событие В – «на кубике выпало нечетное число очков».



Устойчивость частот является скорее математическим, а не экспериментальным фактом. На нем основывается частотное, или ***статистическое определение вероятности:*** за вероятность случайного события можно приближенно принять его относительную частоту.

В теории вероятностей величина, значение которой зависит от исхода эксперимента, называется ***случайной величиной.*** Мы будем рассматривать ряд числовых значений такой величины, полученных в какой-либо серии экспериментов. Такой числовой ряд называется ***случайной выборкой.***

Представим себе, что из всех школьников вашего региона случайным образом выбирается 1000 человек и фиксируется их оценка по математике в последней четверти. В статистике множество всех школьников региона будет называться ***генеральной совокупностью,*** а случайно выбранная 1000 школьников — случайной выборкой. Однако математики предпочитают иметь дело не со школьниками, а с числами, поэтому мы с вами *будем называть случайной выборкой только числовой ряд* — т. е. в нашем случае не самих выбранных **школьников**, а их оценки.

Займемся числовыми рядами, полученными в результате описанной процедуры.

Пример 1. Среди школьников седьмых классов был проведен выборочный опрос: из скольких человек состоят их семьи? В результате такого опроса была получена следующая выборка:

2 2 3 3 3 3 4 2 3 3 2 3 2 3 2 3 2 4 3 2 2 3 2 4 5 2 3 3 2 4 3 2 3 4 3 2 3 5 3

Здесь каждое число означает количество человек в семье соответствующего ученика. Числа выписаны в том порядке, в котором ученики сдавали свои ответы. А что если упорядочить эти числа по возрастанию? 2 2 2 2 2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 5 5

Не правда ли, этот ряд (он называется ***ранжированным***) воспринимается уже лучше: теперь мы ясно видим, что минимальное значение в нем равно 2, а максимальное —5. Видно, как часто повторяется каждое из значений.

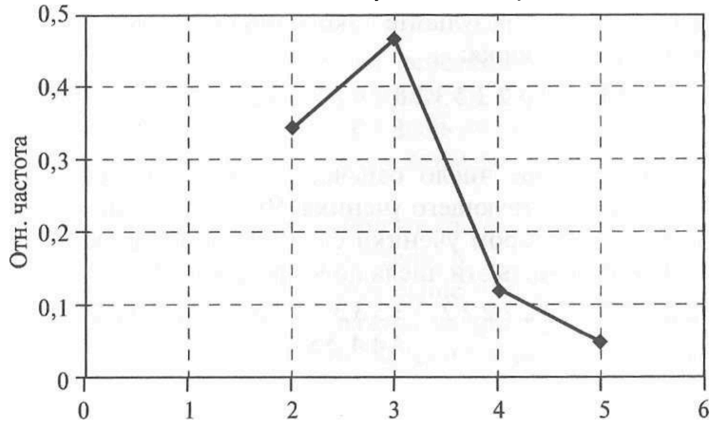
А вот как выглядит представление выборки в ***виде частотной таблицы:***

Состав семьи (количество человек)	Абсолютная частота	Относительная частота
2	14	0,35
3	19	0,475
4	5	0,125
5	2	0,05

Первый столбец частотной таблицы содержит *различные значения* наблюдаемой величины (по возрастанию), второй — сколько раз это значение повторилось в выборке, т. е. его ***абсолютную частоту***, третий — какую долю эти значения составляют от всей выборки, т. е. его ***относительную частоту.***

Разумеется, сумма абсолютных частот будет равна ***объему выборки*** (т. е. количеству опрошенных учеников — 40), а сумма относительных —1.

Еще более наглядной формой представления результатов выборки является их графическое изображение. Для этого используется так называемый ***полигон частот*** — линейная диаграмма, на которой по горизонтальной оси откладываются различные значения, полученные в выборке, а по вертикальной — их относительная частота. После этого полученные точки соединяются ломаной линией (отсюда и название — *полигон* в переводе с греческого означает *многоугольник).*



Количество человек в семье

Пример 2. На школьниках 1 «А» класса было проведено исследование для выяснения того, сколько весит портфель первоклассника. В результате взвешиваний был получен следующий числовой ряд (вес каждого портфеля в кг):
2,1; 2,45; 1,9; 2,6; 3,1; 1,95; 3,4; 4,3; 1,15; 2,7;2,2; 3,2; 2,4; 2,2; 1,8; 1,5; 2,4; 2,25; 2,6; 1,75.

Чем принципиально отличаются результаты этой выборки от примера 1? Чтобы ответить на этот вопрос, попробуем составить для нее таблицу частот:

Вес портфеля (в кг)	Абсолютная частота	Относитель-ная частота
1,15	1	0,05
1,5	1	0,05
1,75	1	0,05
1,8	1	0,05
1,9	1	0,05
1,95	1	0,05
2,1	1	0,05
2,2	2	0,05
2,25	1	0,05
2,4	2	0,1
2,45	1	0,05
2,6	2	0,1
2,7	1	0,05
3,1	1	0,05
3,2	1	0,05
3,4	1	0,05
4,3	1	0,05

Как видите, частота каждого значения оказалась равной 1 или 2. Это неудивительно, ведь точные совпадения в такой выборке маловероятны, а если измерять вес портфелей еще точнее, то совпадений не будет вовсе. Ясно, что составлять для такой выборки таблицу частот или рисовать полигон бессмысленно — никакого наглядного представления мы при этом не получим.

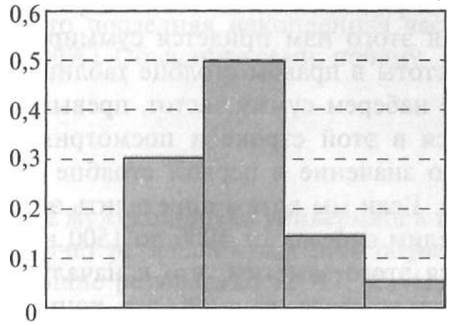
В такой ситуации для представления результатов выборки используют ***интервальную таблицу частот:*** весь диапазон значений выборки разбивают на интервалы (чаще всего равные) и подсчитывают частоту попадания в каждый интервал. Вот как будет выглядеть такая таблица в нашем примере, если разбить диапазон от 1 до 5 кг на четыре равных интервала:

Вес портфеля (в кг)	Абсолютная частота	Относительная частота
от 1 до 2	6	0,3
от 2 до 3	10	0,5
от 3 до 4	3	0,15
от 4 до 5	1	0,05

При попадании значения на границу интервалов его относят к какому-то одному из них (например, левому), чтобы не считать дважды. Так, если бы у кого-то из первоклассников портфель весил ровно 3 кг, мы включили бы это значение в интервал от 2 до 3 кг.

Данные интервальной таблицы частот принято представлять уже не полигоном, а ***гистограммой частот:*** по горизонтальной оси откладываются интервалы значений, а над каждым интервалом строится *столбик, площадь которого равна относитель-*

ной частоте попадания в данный интервал. Обратите внимание: именно площадь, а не высота. Хотя, если интервалы равные, то высоты всех столбиков отличаются от соответствующих частот только постоянным множителем — длиной интервала.



1 2 3 4 5

Вес портфеля, кг

В некоторых задачах таблицу частот удобно дополнить еще одной характеристикой, получившей название ***накопленной частоты.*** Рассмотрим ее использование на примере.

Пример 3. Перед вами еще одна интервальная таблица частот — распределение семей по уровню доходов:

Доход на человека (в руб.)	Относительная частота
менее 500	2%
от 500 до 1000	6%
от 1000 до 1500	7%
от 1500 до 2000	12%
от 2000 до 2500	36%
от 2500 до 3000	27%
свыше 3000	10%

Предположим, вы услышали по телевизору фразу: «Около 12% семей живет сейчас за чертой бедности». Попробуем определить по имеющейся у нас таблице эту «черту». Для этого нам придется суммировать относительные частоты в правом столбце таблицы до тех пор, пока мы не наберем сумму частот, превышающую 12%. Остановимся в этой строке и посмотрим, чему в это время равно значение в первом столбце — от 1000 до 1500 рублей. Если мы хотим определить эту черту более точно, поделим отрезок от 1000 до 1500 в нужной пропорции. Для этого заметим, что к началу этого отрезка сумма частот составляла 8%, а к концу стала равна 15%. Значит, интересующее нас значение *x* можно найти из пропорции:

$$\frac{12\% - 8\%}{15\% - 8\%} = \frac{x - 1000}{1500 - 1000}, \qquad x \approx 1285$$

1285 рублей — это и есть та самая черта, которую диктор назвал «уровнем бедности».

Решая эту задачу, мы должны были производить накопительное суммирование относительных частот до тех пор, пока не будет достигнут заданный уровень — 12%. Поскольку эти результаты можно использовать и для решения других задач, удобно хранить полученные результаты — ***накопленные частоты*** — в отдельном столбце таблицы:

Доход на человека (в р.)	Относительная частота	Накопитель-ная частота
менее 500	2%	2%
от 500 до 1000	6%	8%
от 1000 до 1500	7%	15%
от 1500 до 2000	12%	27%
от 2000 до 2500	36%	63%
от 2500 до 3000	27%	90%
свыше 3000	10%	100%

Отметим, что последняя накопленная частота всегда равна 1 (или 100%).

Используя данные последней таблицы, решим задачи, встречающиеся на ЕГЭ.

Пример 4. Дана указанная таблица, *но без графы* «Относительная частота». Вопрос: с какой вероятностью доход людей колеблется от 2500 руб. до 3000 руб.?

Решение: Вероятность, как относительную частоту мы найдем вычитанием из накопительной частоты, стоящей напротив интересующего нас интервала дохода на человека от 2500р. до 3000р. (90%), предыдущего значения накопительной частоты (63%). Т.е. 90% – 63% = 27% = 0,27.

Ответ 0,27.

Пример 5. Дана указанная таблица. Определить сколько человек из 400 получают наиболее «популярную» зарплату?

Решение: «Популярная» зарплата соответствует строке таблицы напротив частоты 36%. Это значит, из ста человек - 36 получают зарплату от 2500р. до 3000р., тогда из пропорции найдем: из четырехсот человек $4 \cdot 36 = 144$ человека имеют такой же доход.

Ответ 144.