

Способы работы статистических данных

*Ученик группы 43и1
Пиговаев Дмитрий Иванович*

Статистическая обработка данных

Основные статистические
характеристики

Основные этапы статистической обработки данных

Пример

В ходе некоторого анкетирования были получены следующие ответы: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 1, 3, 4, 5, 7, 9, 1, 5, 7, 9, 3, 5, 3, 5, 3, 5, 3, 5, 5, которые занесли в таблицу

Ответ	0	1	2	3	4	5	6	7	8	9	10
Кол-во ответов	2	5	3	9	4	10	3	5	3	5	1

Основные этапы статистической обработки данных

Каждый ответ (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10) называется - **варианта измерения**.

Если все варианты записать по порядку (например, по времени и т.п.) то получится **ряд данных**.

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 1, 2, 3, 4, 5, 6, 7,
8, 9, 1, 3, 4, 5, 7, 9, 1, 5, 7, 9, 3, 5, 3, 5, 3, 5, 3, 5, 5

Если же все варианты записать в порядке неубывания, то получится **сгруппированный ряд данных**.

0, 0, 1, ..., 1, 2, 2, 2, 3, ..., 3, 4, ..., 4, 5, ..., 5, 6, 6, 6, 7, ..., 7, 8, 8, 8, 9, ..., 9,
10

2 5 3 9 4 10 3 5 3 5 1

Если среди всех данных одна из вариантов встретилась k раз, то число k называют **кратностью** этой варианты.

Основные этапы статистической обработки данных

Таким образом, *таблица распределения данных* имеет

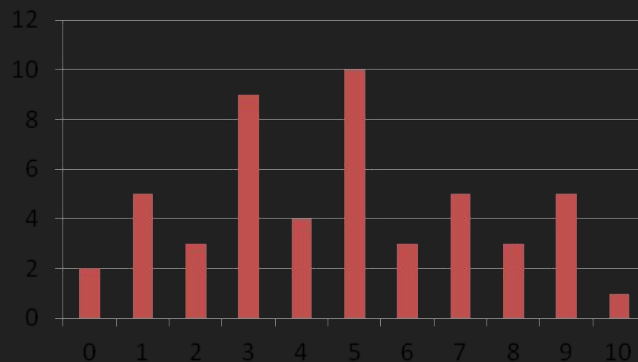
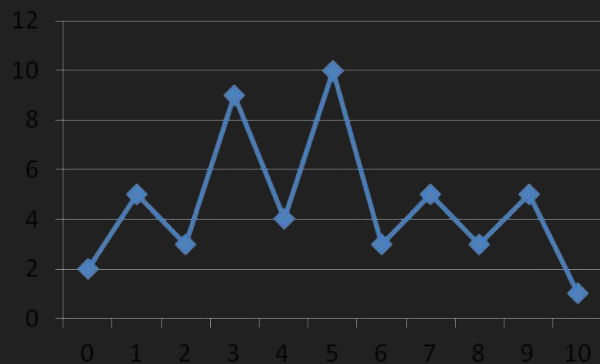
	Варианта											Сумма
	0	1	2	3	4	5	6	7	8	9	10	а
Кратность	2	5	3	9	4	10	3	5	3	5	1	50

Сумма (50) всегда равна сумме кратностей
(2+5+3+9+4+10+3+5+3+5+1)

Основные этапы статистической обработки данных

	Варианта											Сумма
	0	1	2	3	4	5	6	7	8	9	10	а
Кратность	2	5	3	9	4	10	3	5	3	5	1	50

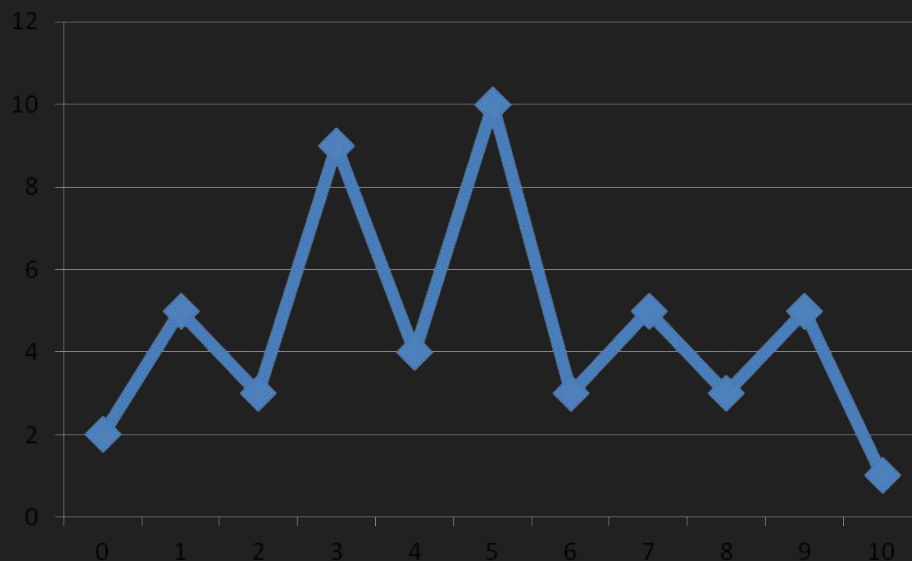
По таблице распределения данных строят три вида диаграмм



Основные этапы статистической обработки данных

	Варианта											Сумм
	0	1	2	3	4	5	6	7	8	9	10	а
Кратн ость	2	5	3	9	4	10	3	5	3	5	1	50

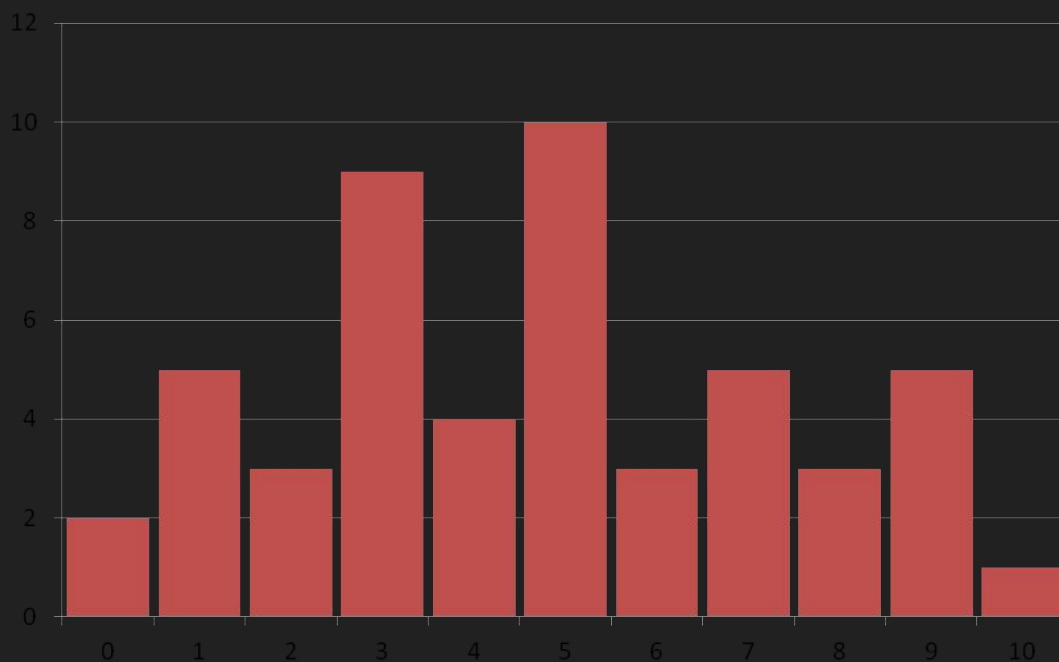
Многоугольник распределения



Основные этапы статистической обработки данных

	Варианта											Сумма
	0	1	2	3	4	5	6	7	8	9	10	а
Кратность	2	5	3	9	4	10	3	5	3	5	1	50

Гистограмма распределения



Основные этапы статистической обработки данных

	Варианта											Сумма
	0	1	2	3	4	5	6	7	8	9	10	
Кратность	2	5	3	9	4	10	3	5	3	5	1	50

Круговая диаграмма



Основные этапы статистической обработки данных

1. Упорядочивание и группировка
2. Составление таблицы распределения данных
3. Построение графиков распределения данных
(многоугольник, гистограмма, круговая диаграмма)
4. Получение паспорта данных измерения
 - Объём измерения
 - Размах измерения
 - Мода измерения
 - Среднее измерения
 - Медиана измерения
 - Частота варианты

меры центральной тенденции

Числовые характеристики измерения

Объём измерения – количество значений измерения.

Размах измерения – разность между наибольшим и наименьшим результатами измерения.

Мода измерения – наиболее часто встречающийся результат измерения.

Среднее измерения – среднее арифметическое всех значений.

Медиана измерения – средняя варианта в сгруппированном ряде данных (если количество значений нечётно) или полусумма двух средних вариантов (если количество значений чётно).

Частота варианты – отношение кратности варианты к объёму измерения; может быть числовой и процентной.

Мода измерения

Числовой характеристикой выборки, как правило, не требующей вычислений, является так называемая мода. Модой называют количественное значение исследуемого признака, наиболее часто встречающееся в выборке. Для симметричных распределений признаков, в том числе для нормального распределения, значение моды совпадает со значениями среднего и медианы. Для других типов распределении, несимметричных, это не характерно. К примеру, в последовательности значений признаков 1, 2, 5, 2, 4, 2, 6, 7, 2 модой является значение 2, так как оно встречается чаще других значений - четыре раза.

Медиана

Медианой называется значение изучаемого признака, которое делит выборку, упорядоченную по величине данного признака, пополам. Справа и слева от медианы в упорядоченном ряду остается по одинаковому количеству признаков.

Например, для выборки 2, 3, 4, 4, 5, 6, 8, 7, 9 медианой будет значение 5, так как слева и справа от него остается по четыре показателя. Если ряд включает в себя четное число признаков, то медианой будет среднее, взятое как полусумма величин двух центральных значений ряда. Для следующего ряда 0, 1, 1, 2, 3, 4, 5, 5, 6, 7 медиана будет равна 3,5.

Дисперсия

Числовую характеристику данных измерения, отвечающую за разброс (рассеивание) данных вокруг их среднего значения, называют **дисперсией** (от лат. *disperses* – рассыпанный, разогнанный, рассеянный) и обозначают буквой D ; число $\sigma = \sqrt{D}$ называют **средним квадратическим отклонением**. Чем меньше дисперсия D или среднее квадратическое отклонение σ , тем плотнее группируются данные измерения вокруг своего среднего значения.

Алгоритм вычисления дисперсии

Для нахождения дисперсии D данных $x_1, x_2, \dots, x_n$ измерения следует вычислить:

- 1) Среднее значение $M = (x_1 + x_2 + \dots + x_n) / n$
 - 2) Отклонения данных от M , т.е. $x_1 - M, x_2 - M, \dots, x_n - M$
 - 3) Квадраты $(x_i - M)^2$ отклонений, найденных на предыдущем шаге
 - 4) Среднее значение всех квадратов отклонений
 $D = [(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2] / n$ это и есть дисперсия.
- σ = Корень из D среднее квадратическое отклонение.

Размах измерения

Размах — разность между наибольшим и наименьшим значениями результатов наблюдений. Пусть $X_1, \dots, X_n$ — взаимно независимые случайные величины с функцией распределения $F(x)$ и плотностью вероятности $f(x)$. В этом случае размах W_n определяется как разность между наибольшим и наименьшим значениями среди $X_1, \dots, X_n$; размах представляет собой случайную величину, которой соответствует функция распределения:

$$P\{W_n \leq \omega\} = n \int_{-\infty}^{\omega} [F(\omega + x) - F(x)]^{n-1} f(x) dx$$

Среднее измерения

Среднее измерения (среднее арифметическое) значение как статистический показатель представляет собой среднюю оценку изучаемого в эксперименте психологического качества. Эта оценка характеризует степень его развития в целом у той группы испытуемых, которая была подвергнута психодиагностическому обследованию. Сравнивая непосредственно средние значения двух или нескольких выборок, мы можем судить об относительной степени развития у людей, составляющих эти выборки, оцениваемого качества.

Спасибо за внимание!