

Первичное описание исходных данных

Для первичного описания исходных данных используются *таблицы и графики*.

Обычно в ходе исследования признак измеряется не у одного - двух, а у множества объектов (испытуемых). Кроме того, каждый объект характеризуется не одним, а целым рядом признаков, измеренных в разных шкалах.

Первичное описание исходных данных

Результаты измерения для дальнейшего анализа чаще всего представляют в виде **таблицы** исходных данных. Каждая строка такой таблицы обычно соответствует одному объекту, а каждый столбец - одному измеренному признаку. Таким образом, исходной формой представления данных является **таблица типа «объект - признак»**.

Таблица данных

Таблица исходных данных

Таблица 1

1	2	3	4	5	
№	Ф.И.О.	Пол	Класс	Самооценка	
				До	После
		1	2	3	4
1	Иванов И. О.	1	0	5	5
2	Васильев К. А.	1	1	3	4
3	Розова М. И.	0	1	2	3
4	Краснова О. С.	0	0	3	3
5	Цветов С. Т.	1	0	1	3
6	Лозовая Е. И.	0	1	4	4
...	...	...	...	...	...
i	...	x_{1i}	x_{2i}	x_{3i}	x_{4i}
...	...	...	...	...	...
60	Петров Е. М.	1	1	3	3

X_{3i} - самооценка до тренинга,

X_{4i} - самооценка после тренинга,

где i - текущий номер испытуемого (меняется от 1 до $N=60$).

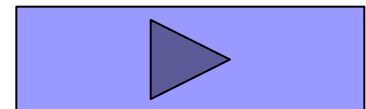


Таблица частот

Частотой события называется количество случаев появления событий.

Для изучения частоты встречаемости значений интересующего признака (переменной) в имеющемся множестве наблюдений строятся *таблицы и графики распределения частот*.

Если указывается, сколько раз встречается каждое значение признака, то это - ***таблица абсолютных частот распределения***, если указывается доля наблюдений, приходящихся на то или иное значение признака, то говорят об ***относительных частотах распределения*** (частность).

Еще одной разновидностью таблиц распределения являются ***таблицы распределения накопленных частот***. Они показывают, как накапливаются частоты по мере возрастания значений признака.

Таблица частот

Пример: Необходимо распределить ответы всех учеников до проведения тренинга. Для этого подсчитывают частоту встречаемости каждого из ответов и составляют таблицу распределения частот (табл. 2).

Таблица показывает, что чаще встречаются средние значения выраженности признака и реже - крайние значения.

Таблица 2

Таблица распределения частот

Значение	f_a (абсолютная частота)	f_o (относительная частота)	f_{cum} (накопленная частота)
5	3	0,05	1,00
4	12	0,20	0,95
3	21	0,35	0,75
2	15	0,25	0,40
1	9	0,15	0,15
Σ (сумма):	60	1	—



Таблица частот

$$f_0 = \frac{f_a}{N}$$

где f_a — абсолютная частота некоторого значения признака, N — число наблюдений, f_0 — относительная частота этого значения признака (частность).

Очевидно, что сумма всех абсолютных частот равна числу наблюдений - N , а сумма всех относительных частот равна 1. Нередко относительная частота применяется для оценки вероятности встречаемости значения.

Таблица сгруппированных частот

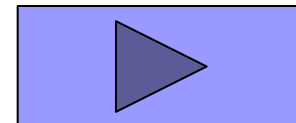
Во многих случаях признак может принимать множество различных значений, например, если мы измеряем время решения тестовой задачи.

В этом случае о распределении признака позволяет судить *таблица сгруппированных частот*, в которых частоты группируются по разрядам или интервалам значений признака.

Таблица 3

**Таблица частот, сгруппированных по интервалам времени решения
тестовой задачи**

Интервал времени, с	f_{aj} (абсолютная частота)	f_{oj} (относительная частота)	f_{cum} (накопленная частота)
30–34	1	0,025	0,025
35–39	2	0,050	0,075
40–44	5	0,125	0,200
45–49	8	0,200	0,400
50–54	10	0,250	0,650
55–59	8	0,200	0,850
60–64	4	0,100	0,950
65–69	2	0,050	1,000
Σ (сумма):	40	1,000	—



Вариационный ряд - упорядоченное отражение распределения значений признака, представляющий двойной ряд чисел и состоящий из обозначения классов и соответствующих частот (частота встречаемости обозначается символом **f**).

Числовой ряд является **непрерывным**.

Таблица – Результаты исследования интроверсии

уровень интроверсии	10	11	12	13	14
f (количество случаев)	10	24	30	0	8

Основные формулы для таблиц сгруппированных частот

1. Число разрядов (количество интервалов)- от 6 до 15: $k = \sqrt{n}$
2. Объем разряда (интервал разряда): $V_{\text{раз}} \geq (R+1)/k$, где R-размах, k – количество разрядов.
3. Количество «лишних вариантов (заступов)» =
 $= V_{\text{раз. округ}} * k - R - 1$.

Задание: в группе испытуемых численностью 42 человек измерено время решения тестовой задачи. Были получены следующие значения: {63,35,58,53,45,61,37,58,51,40,60,55,43,57,50,61,44,64,45,77,53,46,47,72,49,51,64,52,56,58,59,60,49,64,32,51,64,65,69,66,40,69}.

Постройте таблицу сгруппированных частот.

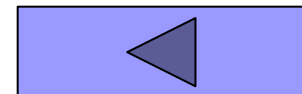
Таблицы сопряженности, или кросстабуляции

Таблицы сопряженности, или кросстабуляции - это таблицы совместного распределения частот двух и более номинативных признаков, измеренных на одной группе объектов. Эти таблицы позволяют сопоставить два или более распределения. Столбцы такой таблицы соответствуют категориям (градациям) одного номинативного признака, а строки — категориям (градациям) другого номинативного признака.

Зависимость распределения оставленных и полученных открыток
от их содержания

		Сообщения	
		отправлено	не отправлено
Новость	Хорошая	35	25
	Плохая	23	97
Σ		58	122

ЗАДАНИЕ: Построить таблицу сопряженности признаков «Пол» (две градации) и «Самооценка» (пять градаций) по таблице 1.





Графики - это чертежи, которые можно использовать для наглядности распределения количественно выраженной величины в выборке.

Диаграммы используются главным образом для изображения соотношений между величинами. Это способ графического изображения величин при помощи фигур (секторов, столбцов и т.п.), площади которых пропорциональны этим величинам.

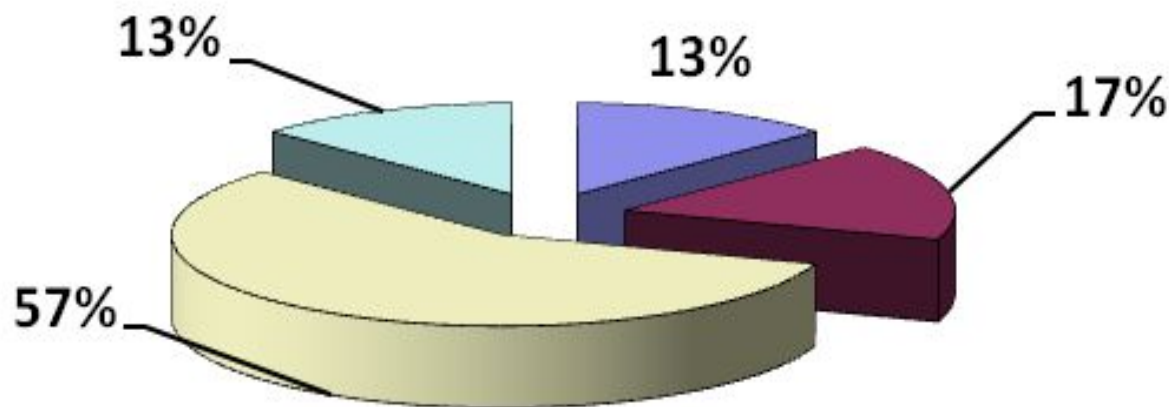
.



В зависимости от круга решаемых задач выделяют диаграммы сравнения, структурные диаграммы, диаграммы динамики.

Особым видом графиков являются диаграммы распределения величин, представленных вариационным рядом. Это гистограмма, полигон и др.

Секторная диаграмма – диаграмма, в которой числа (обычно проценты) изображены в виде круговых секторов,



Примечание: ■ ЭГ-1 ■ ЭГ-2 ■ ЭГ-3 ■ ЭГ-4

Рисунок 1- Секторная диаграмма

Графики (гистограмма распределения частот)

Для более наглядного представления строится *график распределения частот* или *график накопленных частот*.

Гистограмма распределения частот - это столбчатая диаграмма, каждый столбец которой опирается на конкретное значение признака или разрядный интервал (для сгруппированных частот). Высота столбика пропорциональна частоте встречаемости соответствующего значения.

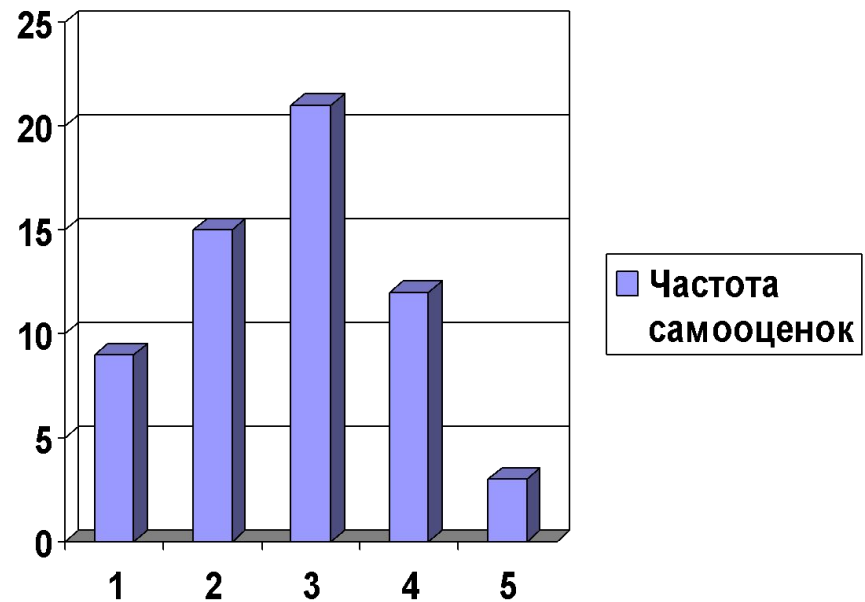
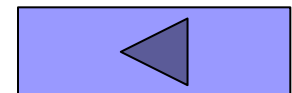


Рис. 2. Гистограмма распределения частот самооценки (по данным таблицы 2)



Графики (гистограмма накопленных частот)

Гистограмма накопленных частот отличается от гистограммы распределения тем, что высота каждого столбика пропорциональна частоте, накопленной к данному значению (интервалу). На рис. 3 изображена гистограмма накопленных частот для данных табл. 2.

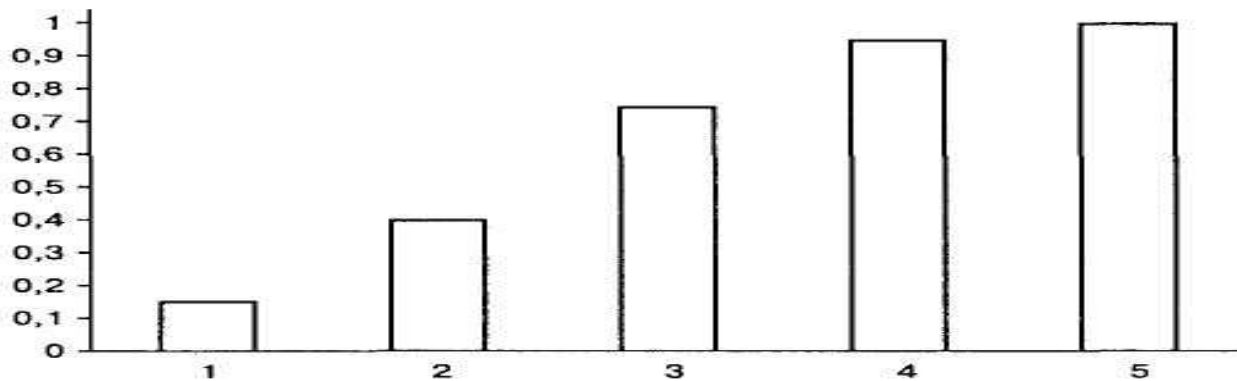
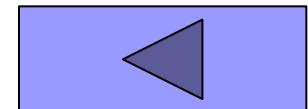


Рис. 3. Гистограмма накопленных относительных частот самооценки



Полигон распределения частот

Построение полигона частот напоминает построение гистограммы. В гистограмме вершина каждого столбца, соответствующая частоте встречаемости данного значения (интервала) признака, - отрезок прямой.

А для полигона отмечается точка, соответствующая середине этого отрезка. Далее все точки соединяются ломаной линией (рис. 4).

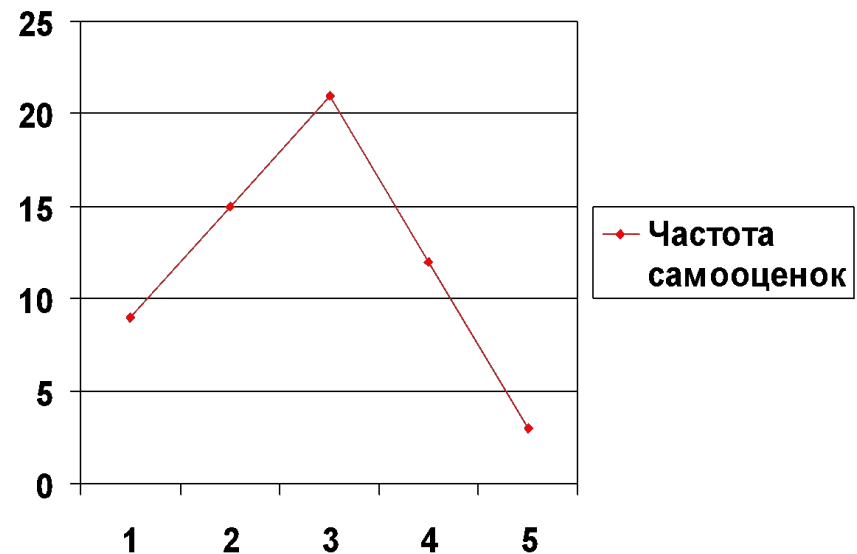


Рис. 4. Полигон распределения частот самооценки (по данным таблицы 2)

Сглаженная кривая распределения частот

Вместо гистограммы или полигона часто изображают **сглаженную кривую распределения частот**. На рис. 5 изображена гистограмма распределения для примера из табл. 3 (столбики) и сглаженная кривая того же распределения частот.

ГЛАВА 3. ТАБЛИЦЫ И ГРАФИКИ

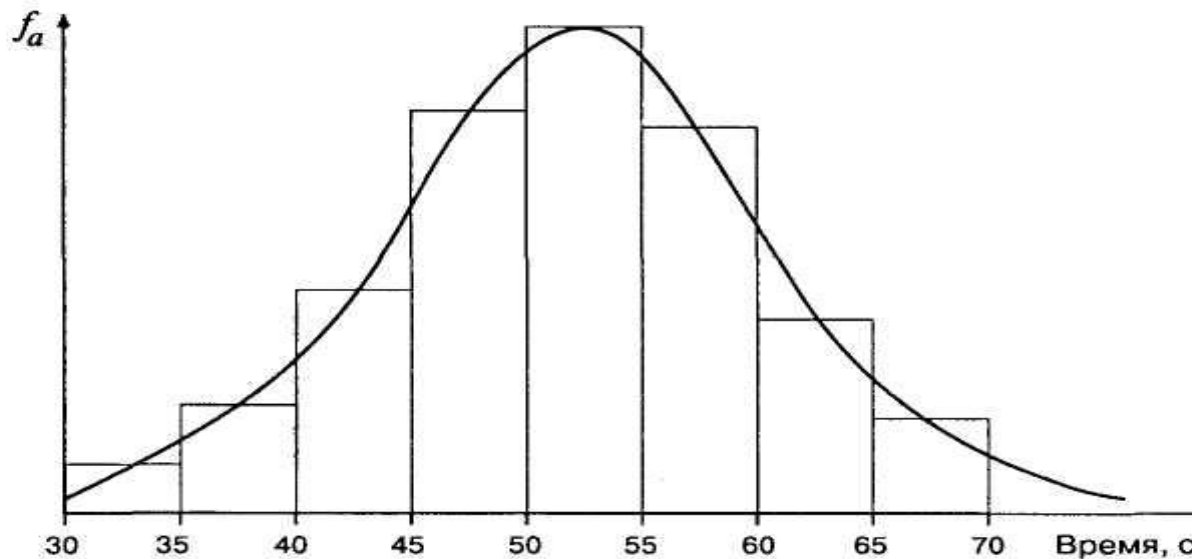
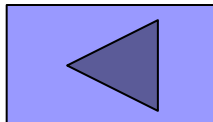


Рис. 3.4. Гистограмма и сглаженный график распределения частот времени решения тестовой задачи (по данным табл. 3.3)



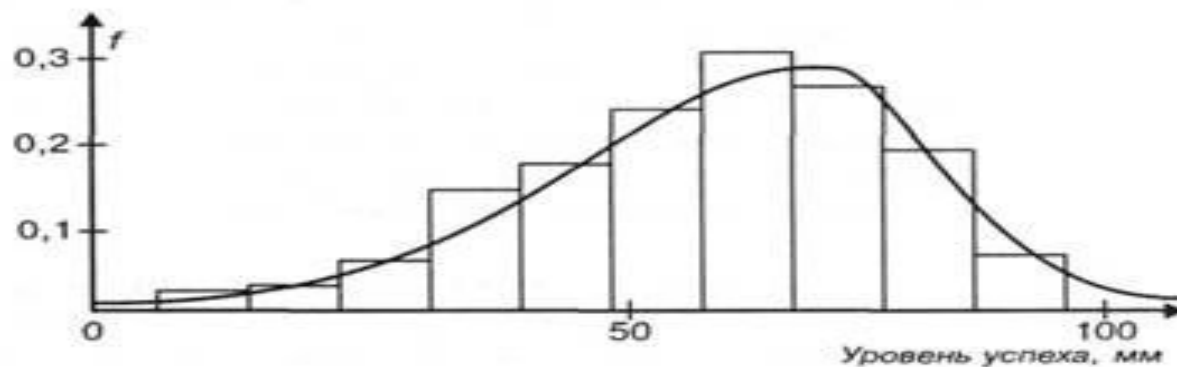


Рисунок 6 – Гистограмма и плавная кривая распределения показателя должного успеха ($n = 356$) (по Е. В. Сидоренко 1997 г.)

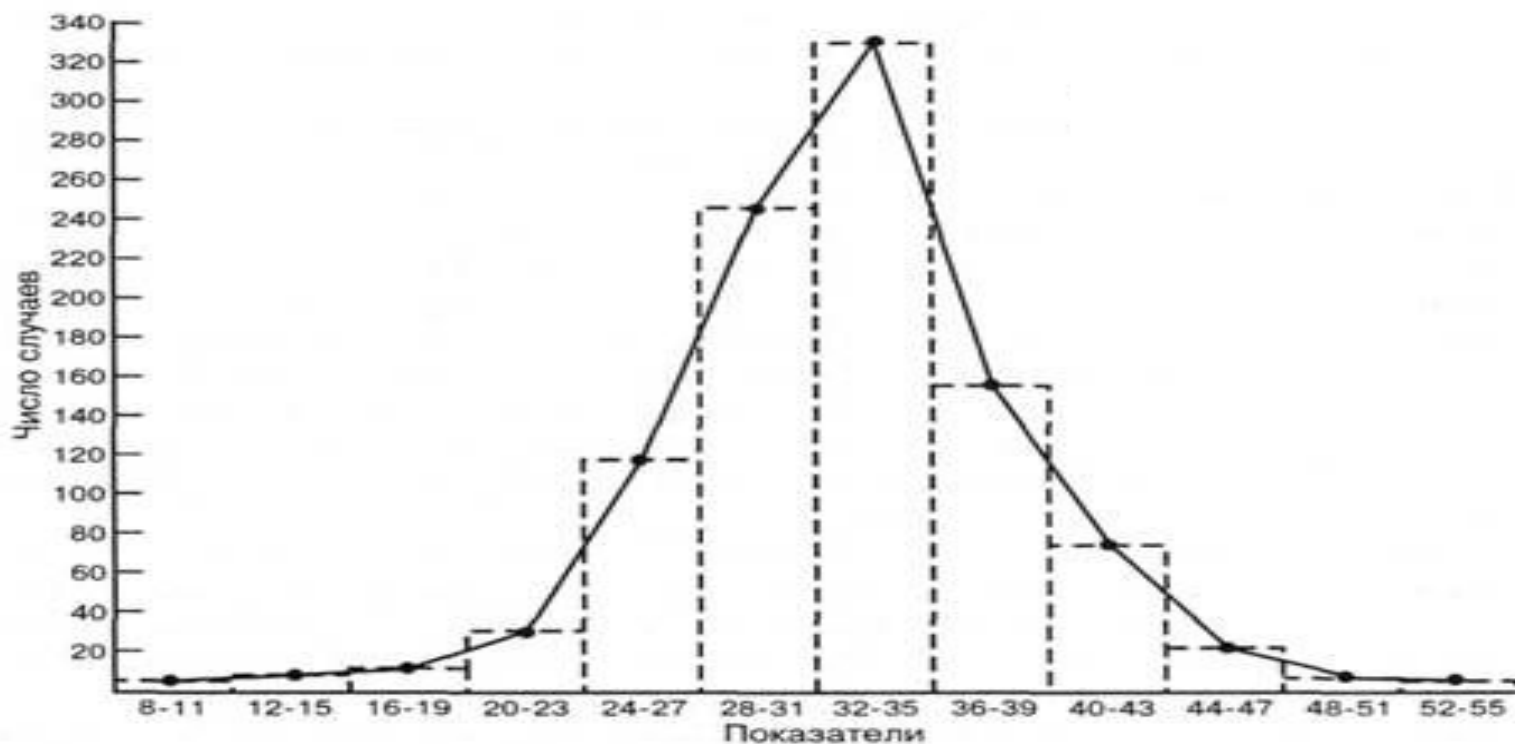
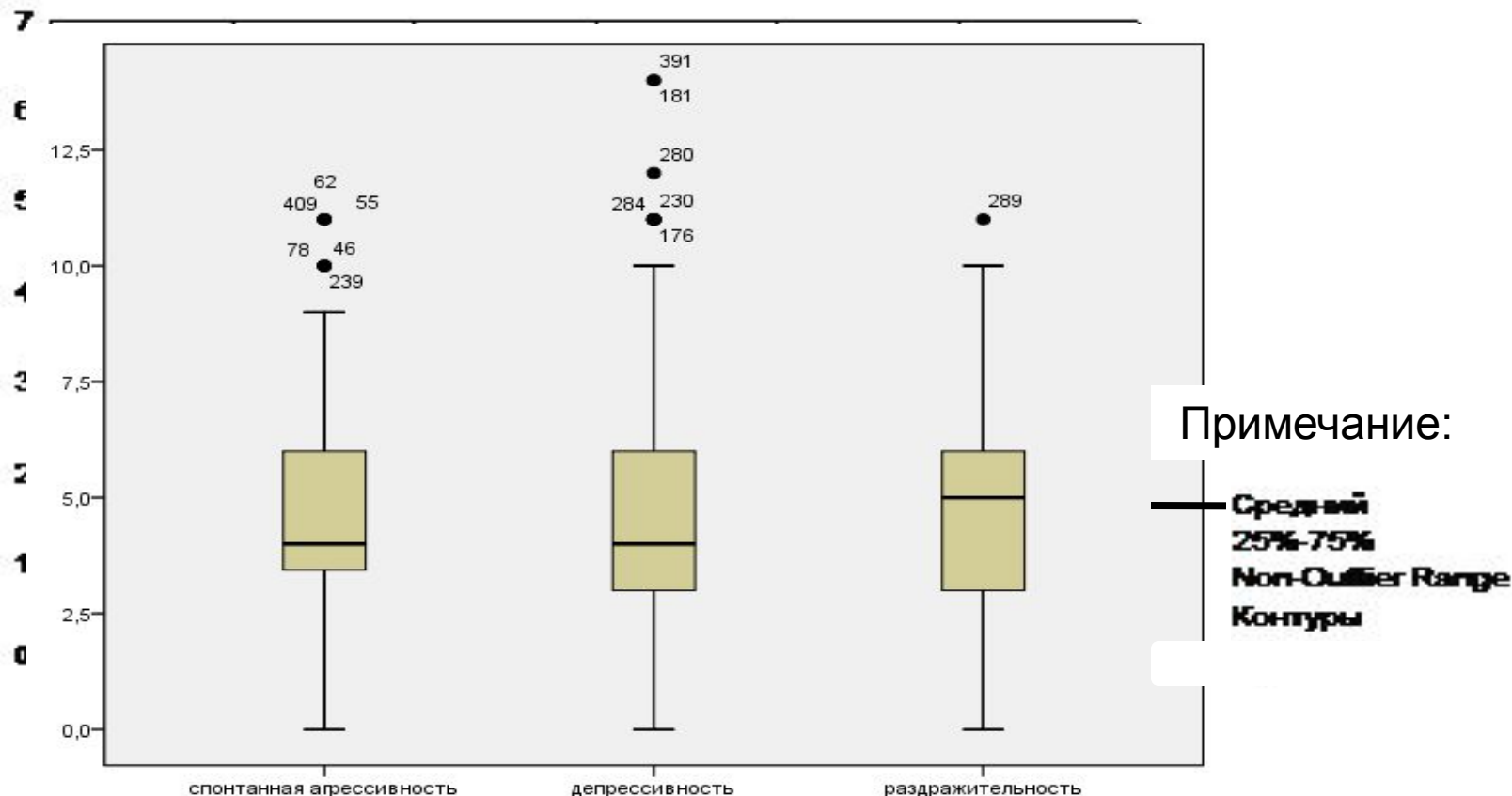


Рисунок 7 – Гистограмма и полигон распределения. Кривая полигона частот и гистограмма

Диаграмма размаха (ящечковая диаграмма) – диаграмма, которая Для каждой группы вычисляется статистика центра (медиана или среднее) и статистики диапазона (например, квартили или стандартные отклонения), и выбранные значения изображаются на диаграмме размаха. Также могут быть изображены точки выбросов (outliers).



Графики следует использовать, когда надо отобразить общий характер функциональной зависимости.

Рекомендации по построению графиков:

- 1) должны включать все необходимые обозначения, чтобы быть понятными сами по себе;
- 2) график и текст должны взаимно дополнять друг друга;
- 3) на одном графике, как правило, не должно быть больше четырех кривых во избежание неразберихи;
- 4) надписи на осях графиков следует располагать внизу и слева;
- 5) линии на графике должны быть разной толщины в зависимости от их важности;
- 6) для обозначения точек наблюдения используются геометрические фигуры.