

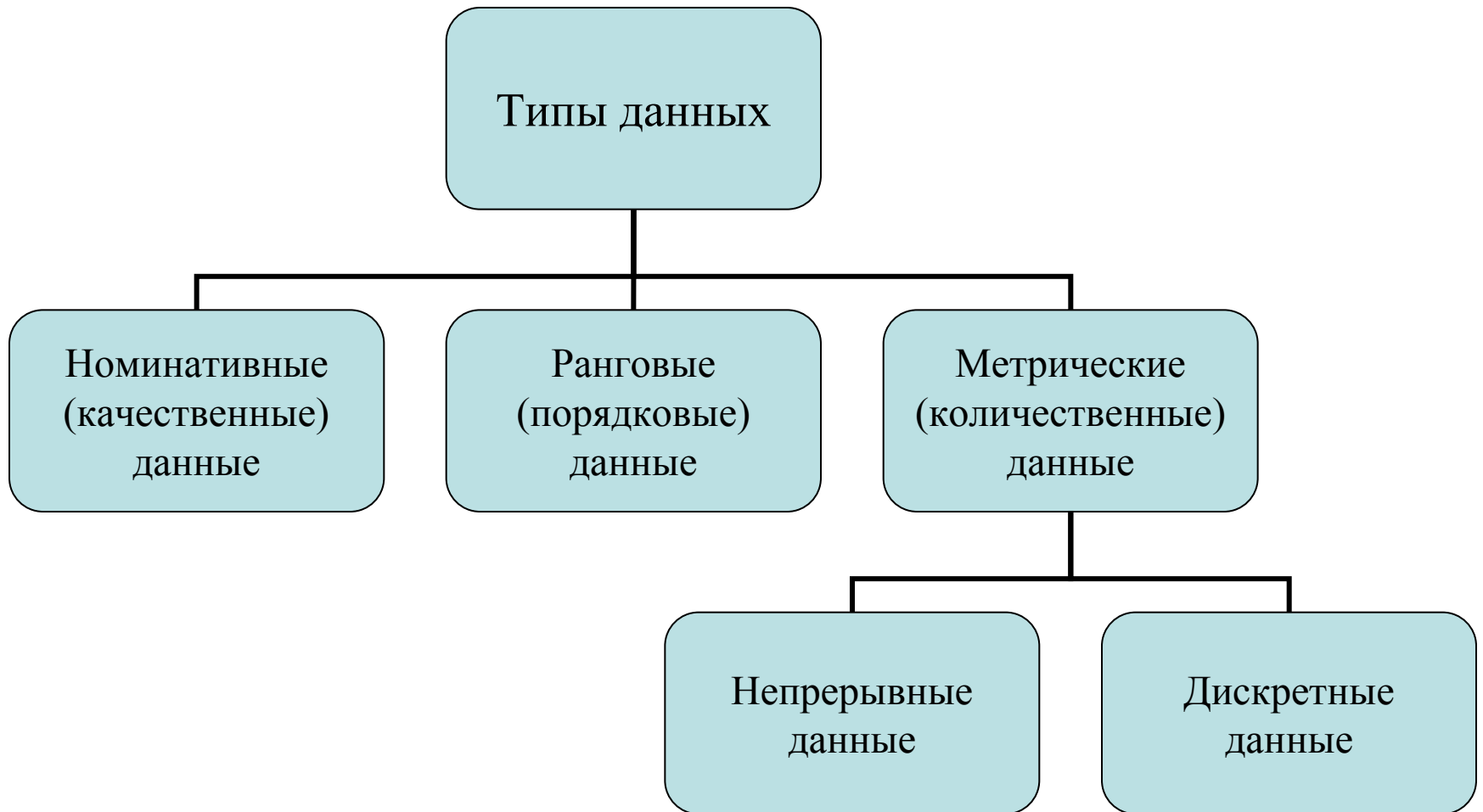
Качественные и количественные методы психологических и педагогических исследований

Тема 3.

Способы представления данных в ПСИХОЛОГИИ

- Представление данных.
- Понятие о квантилях.
- Понятие о рангах. Процедура ранжирования.
- Табулирование данных.
- Графическое представление данных.

Типы данных



Наглядное представление данных

**Наглядное
представление
данных**

```
graph TD; A[Наглядное представление данных] --- B[Табличные данные]; A --- C[Графическое представление данных];
```

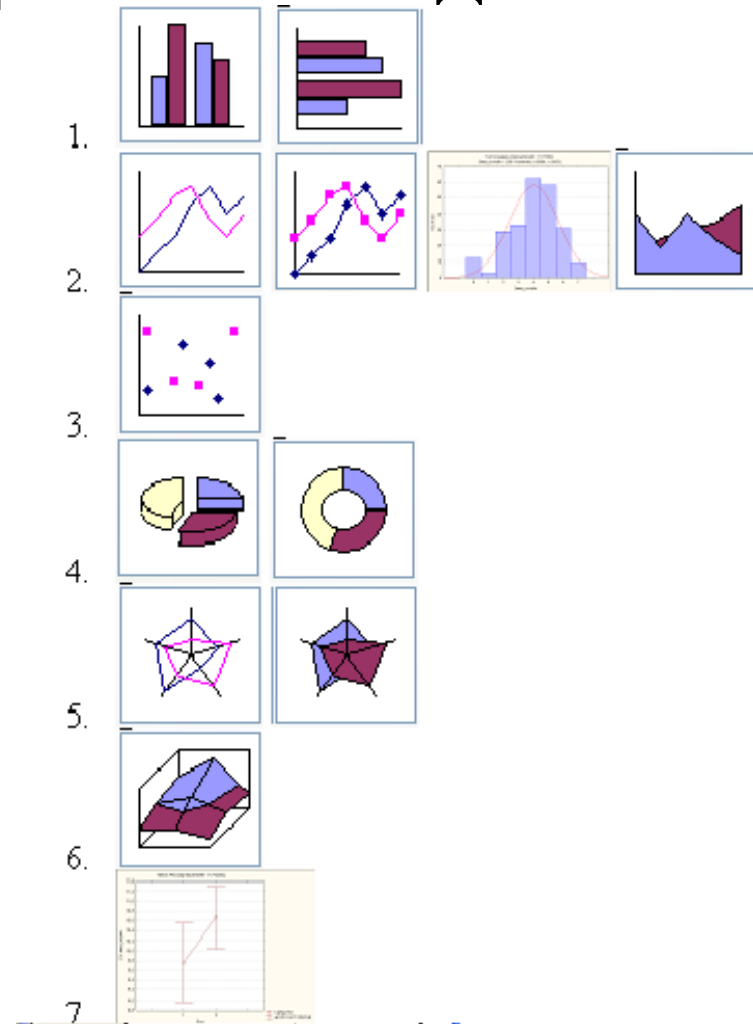
Табличные данные

**Графическое
представление данных**

Графическое представление данных

В самом общем виде диаграммы делятся на:

1. Столбиковые:
 - Вертикальные;
 - Горизонтальные;
2. Линейные
 - Собственно линейные,
 - Ступенчатые,
 - Линейные с областями (профили);
3. Точечные (диаграммы рассеяния);
4. Круговые:
 - Собственно круговая,
 - Кольцевая,
5. Радиальные:
 - Звезды;
 - Лучевые;
6. Диаграммы поверхностей.
7. Комбинированные и др.



Правила графического оформления

- Вся структура графика предполагает его чтение слева направо, вертикальные шкалы — снизу вверх.
- Чтобы диаграмма не получилась сплющенной или вытянутой, выбирают такой масштаб шкалы, чтобы соотношение высоты к ширине составляли 3 к 5.
- На вертикальной шкале необходимо разместить нулевую отметку.
- Пороговые точки на шкалах желательно выделить размером или цветом, но если речь идет о временном интервале, предпочтительно не указывать начальной и конечной точек.
- Подобрать такой масштаб, чтобы кривые линии резко отличались от прямых, желательно включить в график цифровые данные и изображение формулы, а при необходимости — использовать ясные, полные заголовки и подзаголовки как для самой диаграммы, так и для ее осей.

Правила табличного представления первичных данных

- Вся структура таблицы предполагает ее чтение слева направо.
- В первом столбце предполагается размещение испытуемых.
- В последующих столбцах располагаются значения по признакам, полученные после проведения психодиагностической процедуры.

Табулирование данных - это методы и способы построения таблиц

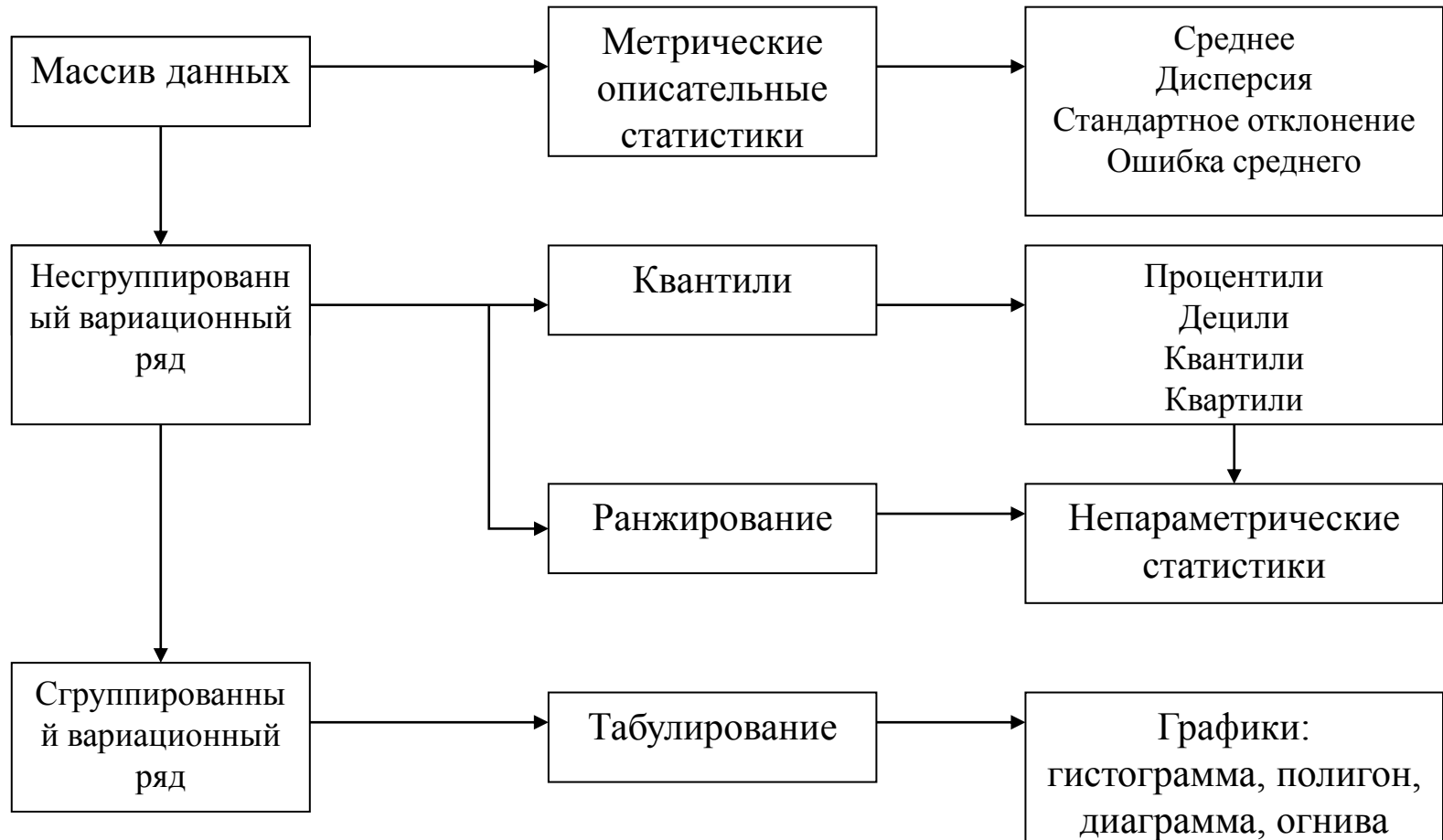
Таблица 1 – Результаты исследования младших школьников

ФИО	Пол	Тревож ность	Идент ичнос ть	Мотивац ия	Усп евае мос ть
МИО	М	3	0	10	3
ВПР	Ж	3	1	20	5
СМТ	Ж	0	0	15	4
ВЛР	М	3	0	12	3
ЖДО	М	5	1	25	5
СТВ	М	0	1	13	3
МИН	М	4	0	18	4
КГН	М	3	1	14	3

Представление данных в психологии бывает в виде:

- **Массив данных** – первичные результаты измерения искомых параметров сводятся в одну таблицу.
- **Несгруппированный вариационный ряд** – упорядочение всех значений переменной от минимального до максимального.
- **Сгруппированный вариационный ряд** – вариационный ряд сворачивают, указывая все полученные значения однократно, а в соседнем столбце указывают частоту, с которой встречается данная оценка

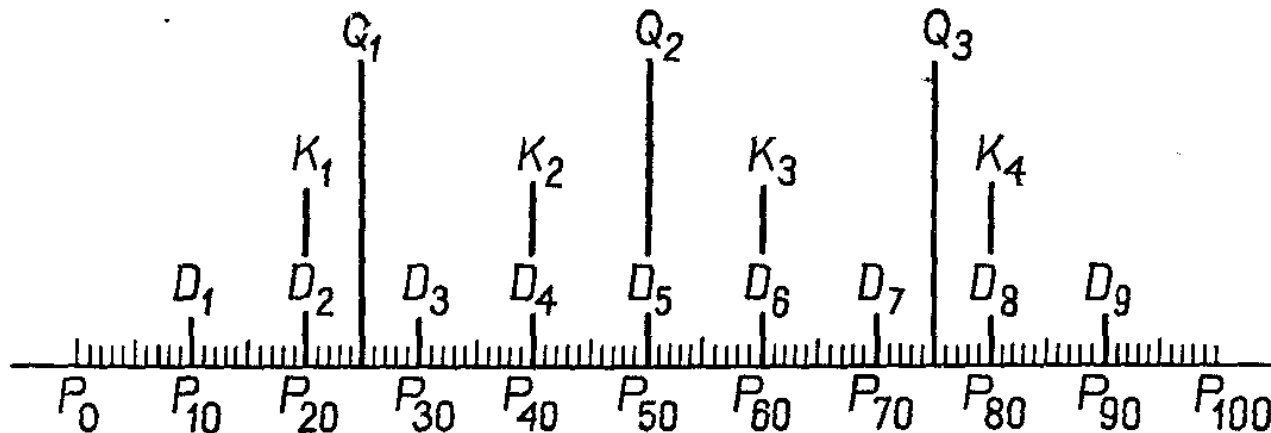
Варианты представления данных



Меры положения – квантили

Квантиль — это точка на числовой оси измеренного признака, которая делит всю совокупность упорядоченных измерений на две группы с известным соотношением их численности

- **Процентиль** (*Percentiles*) — это 99 точек — значений признака ($P_1 \dots, P_{99}$), которые делят упорядоченное (по возрастанию) множество наблюдений на 100 частей, равных по численности.
- **Дециль** - это 9 точек — значений признака ($D_1 \dots, D_9$), которые делят упорядоченное (по возрастанию) множество наблюдений на 10 частей, равных по численности.
- **Квинтель** - это 4 точки — значений признака ($K_1 \dots, K_4$), которые делят упорядоченное (по возрастанию) множество наблюдений на 5 частей, равных по численности.
- **Квартиль** - это 3 точки — значений признака ($Q_1 \dots, Q_3$), которые делят упорядоченное (по возрастанию) множество наблюдений на 4 части, равных по численности.



Нахождение процентиля

- *Процентили указывают на относительное положение индивида в выборке стандартизации.*
- ***P*-й перцентиль** представляет собой точку, ниже которой лежит *P* % процентов всех наблюдений.

Формула

$$P_p = L + \frac{pn - (\text{cum } f)}{f},$$

где **L** – фактически нижняя граница единичного интервала оценок, содержащего частоту **pn**;

cum f - накопленная к **L** частота (до данного интервала);

f – частота оценок в интервале, содержащем частоту **pn**

Задача: Преподаватель предложил 125 учащимся контрольное задание, состоящее из 40 вопросов. В качестве оценки теста выбиралось количество вопросов, на которые были получены правильные ответы. Найти 25-й процентиль

Нахождение интервала:

- Найти между какими значениями в разряде оценок лежит накопленная рп частота (31,25 лежит между 28 и 29 значениями).
- Определить сколько единиц составляет интервал, и разделить пополам (между 28 и 29 лежит $1 / 2 = 0,5$).
- Прибавить к каждому значению интервала результат второго шага ($28 + 0,5 = 28,5$ и $29 + 0,5 = 29,5$)
- Таким образом, искомый интервал лежит между 28,5 и 29,5, а его фактически нижняя граница составляет $L = 28,5$.

Оценка в тесте	Частота	Накопленная частота	Вычисления
38	1	125	<p>Шаг 1. $0,25 n = \frac{n}{4} = \frac{125}{4} = 31,25$</p> <p>Шаг 2. Найти фактическую нижнюю границу разряда оценок, содержащего оценку 31,25: $L = 28,5$</p> <p>Шаг 3. Вычесть накопленную к L частоту из 31,25 $31,25 - 16 = 15,25$</p> <p>Шаг 4. Разделить результат 3-го шага на частоту f в интервале, содержащем оценку 31,25 $\frac{15,25}{18} = 0,85$</p> <p>Шаг 5. Прибавить результат 4-го шага к L $P_{25} = 28,5 + 0,85 = 29,35$</p>
37	1	124	
36	3	123	
35	5	120	
34	9	115	
33	8	106	
32	17	98	
31	23	81	
30	24	58	
29	18	34	
28	10	16	
27	3	6	
26	1	3	
25	0	2	
24	2	2	
	$n = 125$		

сделана запись

Ранговый порядок

Ранжирование – это приписывание объектам чисел в зависимости от степени выраженности измеряемого свойства

- Установите для себя и запомните порядок ранжирования. Вы можете ранжировать испытуемых по их «месту в группе»: ранг 1 присваивается тому, у которого наименьшая выраженность признака, и далее — увеличение ранга по мере увеличения уровня признака. Или можно ранг 1 присваивать тому, у которого 1-е место по выраженности данного признака (например, «самый быстрый»). Строгих правил выбора здесь нет, но важно помнить, в каком направлении производилось ранжирование.
- Соблюдайте правило ранжирования для связанных рангов, когда двое или более испытуемых имеют одинаковую выраженность измеряемого свойства. В этом случае таким испытуемым присваивается один и тот же, средний ранг. Например, если вы ранжируете испытуемых по «месту в группе» и двое имеют одинаковые самые высокие исходные оценки, то обоим присваивается средний ранг 1,5: $(1+2)/2 = 1,5$. Следующему за этой парой испытуемому присваивается ранг 3, и т. д.

Ранжирование данных

Варианты ранжирования

Метрические данные (X_i)	Ранги (X_r)		Метрические данные (X_i)	Ранги (X_r)
15	1	← Принцип: большему значению меньший ранг Принцип: меньшему значению большой ранг →	15	7
11	2		11	6
9	3		9	5
8	4		8	4
7	5		7	3
6	6		6	2
2	7		2	1

Ранжирование связанных рангов

Метрические данные (X_i)	Предварительное ранжирование (X_{R1})	Окончательное ранжирование (X_r)
12	1	1
9	2	$(2+3)/2 = 2,5$
9	3	$(2+3)/2 = 2,5$
7	4	4
6	5	5
5	6	$(6+7+8)/3 = 7$
5	7	$(6+7+8)/3 = 7$
5	8	$(6+7+8)/3 = 7$
4	9	9
2	10	10

Распределение частот

- **Абсолютная частота распределения (f_a)** - называется частота, указывающая, сколько раз встречается каждое значение
- **Относительная частотах распределения (f_o)** – называется частота, указывающая долю наблюдений, приходящихся на то или иное значение признака ($f_o = f_a / N$)
- **Накопленная частота (f_{cum})** – это частота показывающая, как накапливаются частоты по мере возрастания значений признака.
- **Сгруппированная частота** – это частота сгруппированная по разрядам или интервалам значений признака.

Таблица распределения частот

Значени е	f_a (абсолютная частота)	f_o (относительная частота)	f_{cum} (накопленная частота)
5	3	0,05	0,05
4	12	0,20	0,25
3	21	0,35	0,60
2	15	0,25	0,85
1	9	0,15	1
Σ сумма):	60	1	—

Абсолютная и относительная частоты связаны соотношением:

где f_a — абсолютная частота некоторого значения признака,

N — число наблюдений,

f_o — относительная частота этого значения признака.

$$f_o = \frac{f_a}{N},$$

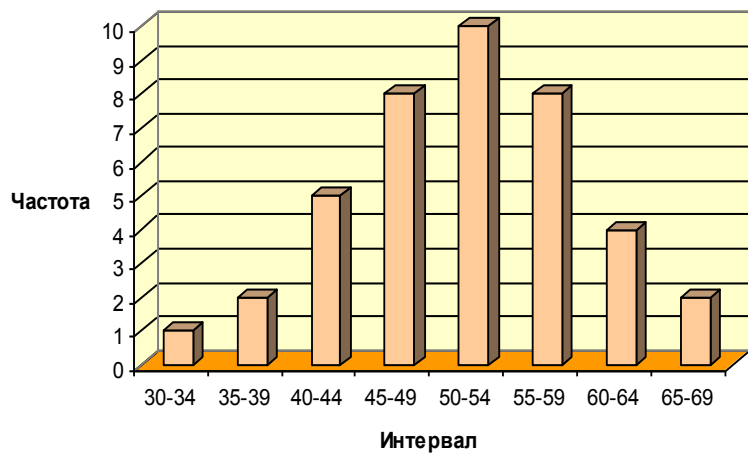
Этапы построения распределения сгруппированных частот

- Уточнение лимитов (крайних значений интервала) – производится округление лимитов - \min и \max значений: реальные лимиты $\max = 67$ и $\min = 32$, уточненные лимиты $\max = 70$ и $\min = 30$.
- Определение размаха: $\max - \min = 70 - 30 = 40$
- Выбор желаемой ширины интервала разрядов l - наиболее удобной шириной интервала разрядов является $l = 5$.
- Определение числа разрядов. Размах делится на интервал разряда: $40/5 = 8$, получаем число разрядов — 8.
- Расчет границ интервалов, посредством прибавления к нижней границе ширину интервала.
- Подсчет абсолютной, относительной и накопленной частот

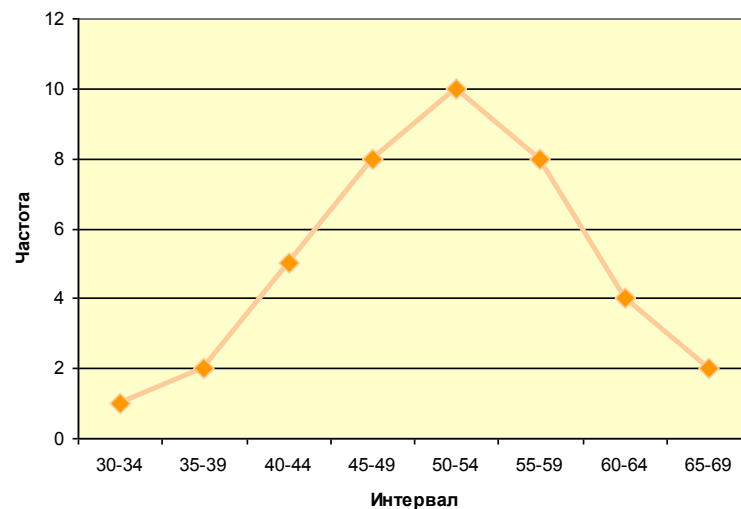
Графическое представление

- **Гистограмма** – это последовательность столбцов, каждый из которых опирается на один отдельный интервал, а высота столбца отражает количество случаев.
- **Вариационная кривая** – линия соединяющая точки, соответствующие середине каждого разрядного интервала и частоте.
- **Полигон распределения** – вариационная кривая с перпендикуляром линий до горизонтальной оси в середине каждого интервала.
- **Полигон накопленных частот (кумулята)** – на оси ординат откладывают значения суммы всех случаев лежащих в данном интервале, так и всех предыдущих интервалов. Сглаженная линия описывает все эти значения.
- **Огива (процентильная кривая)** – сглаженная линия, у которой по оси абсцисс (x) откладывают значения процентов (процентилей), а на оси ординат (y) – значения показателей.
- **Диаграмма** – отражение в долевым отношении частот на круге.

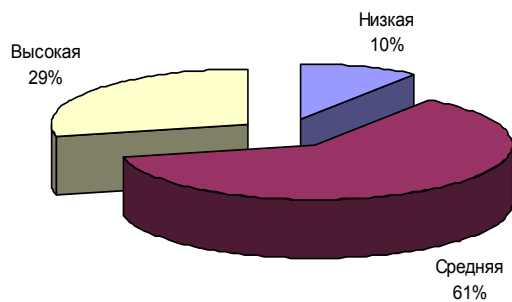
Гистограмма



Полигон распределения частот



Круговая диаграмма



Кумулята

