

Пример: произведем группировку совокупности, включающей 30 студентов одой из групп ЮУрГУ ф-та ЭиП:

№	Рост студента, см	№	Рост студента, см
1	165	16	160
2	172	17	185
3	185	18	173
4	167	19	171
5	166	20	186
6	174	21	189
7	170	22	192
8	160	23	183
9	172	24	187
10	178	25	167
11	168	26	165
12	165	27	161
13	185	28	178
14	184	29	180
15	179	30	179

Вторичная группировка — образование новых групп на основе ранее осуществленной группировки.

1. Определяем n (количество групп группировки) с помощью формулы Стерджесса:

$$n = 1 + 3,322 * \log 30 = 5,906 = 6;$$

2. Рассчитываем шаг (длину интервала группировки):

$$h = (X_{\max} - X_{\min})/n, \text{ где}$$

X_{\max} – максимальное значение признака,

X_{\min} – минимальное значение признака, т.е.

$$h = (192 - 160)/6 = 5,3$$

3. Группируем данные:

Распределение студентов в группы по росту	Число студентов в группе	№ студента, входящего в группу
160-165,3	6	1,8,12,16,26,27
165,3-170,6	5	4,5,7,11,25
170,6-175,9	5	2,6,9,18,19
175,9-181,2	5	10,15,28,29,30
181,2-186,5	6	3,13,14,17,20,23
186,5- 191,8 (192)	3	21,22,24

Проверка: $6+5+5+5+6+3=30$ (должно совпадать с N)

Лекция 4: Статистические ряды распределения

Несистематизированные данные, собранные в процессе стат. наблюдения, образуют **первичный ряд данных**. При достаточно большом объеме совокупности первичный ряд данных становится *трудно обозримым* и непосредственное рассмотрение его *не может дать представления о распределении единиц совокупности по величине признака*.

Первым шагом в упорядочении первичного ряда является его **ранжирование**, то есть расположение всех вариантов ряда (значений признака) *в возрастающем или убывающем порядке*.

Ранжирование данных позволяет:

1) сразу увидеть максимальное и минимальное значение признака в совокупности и оценить разницу между ними ($X_{max} - X_{min}$)

2) определить число повторений отдельных вариантов ряда (частоту).

В результате первичный неупорядоченный ряд данных преобразовывается в упорядоченный ряд, в котором будет отражено число повторений каждой варианты:

$$\begin{array}{c} X_1, X_2, X_3, \dots, X_i, \dots, X_n \\ f_1, f_2, f_3, \dots, f_i, \dots, f_n \end{array}$$

Например, стаж работы (годы) 22 рабочих бригады характеризуется следующими данными:

2, 4, 5, 5, 6, 6, 5, 6, 6, 7, 7, 8, 8, 9, 10, 11, 4, 3, 3, 4, 4, 5

Ранжированный ряд:

2, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 8, 8, 9, 10, 11.

Этот ряд называется **статистическим рядом распределения**. Он характеризует структуру изучаемого явления, позволяет судить о степени однородности изучаемой совокупности, закономерности и границах варьирования анализируемого признака.

Элементами статистического ряда распределения являются варианты — X_i и частоты f_i (абсолютная величина числа повторений i -й варианты).

Для характеристики структуры совокупности используется показатель, который называется **частотью** и определяется по формуле

$$d_i = \frac{f_i}{\Sigma f_i}$$

Из определения частоты и частости следуют равенства:

$$\sum f_i = N; \sum d_i = 1 (100\%), \text{ где } N - \text{объем совокупности.}$$

Ряды распределения могут быть атрибутивными и вариационными.

Атрибутивным называется статистический ряд распределения, построенный по атрибутивному признаку. *Примером* атрибутивных рядов могут служить распределения населения по занятости, национальности, профессии и т.д.

Таблица – Распределение рабочих цеха по профессиям

Профессия рабочего	Количество рабочих (частота)	Удельный вес в общей численности рабочих (частоты)	
		в долях	в %
Токарь	5	0,5	50
Фрезеровщик	2	0,2	20
Слесарь	3	0,3	30
Итого	10	1,0	100

Вариационным называется статистический ряд распределения, построенный по количественному признаку. Вариационный ряд является дискретным, если признак, по которому он построен — дискретный, и интервальным, если признак, по которому он построен — непрерывный.

Примером дискретного ряда может служить распределение рабочих предприятия по квалификации:

Таблица - Распределение рабочих цеха по квалификации

Тарифный разряд рабочего	Количество рабочих (частота)	Удельный вес в общей численности рабочих (частотность)	
		В ДОЛЯХ	В %
2	1	0,05	5
3	5	0,25	25
4	8	0,40	40
5	4	0,20	20
6	2	0,10	10
Итого	20	1,00	100

Данные табл. характеризуют структуру рабочих по квалификации. Наличие такой информации за несколько периодов дает возможность судить о динамике этой структуры, позволяя оценивать изменение качества рабочей силы.

Примером интервального ряда распределения является распределение предприятий *по объему произведенной продукции*, выполненное в процессе построения соответствующей аналитической группировки.

Таблица - Распределение предприятий по объему произведенной продукции

^{xi} Группа предприятия по объему произведенной продукции, млн. руб.	Количество предприятий (частота)	Удельный вес в общей численности предприятий (частотность)		Накопле нная частота
		в долях	в %	
5 – 6,6	3	0,12	12	3
6,6 – 8,2	4	0,16	16	7
8,2 – 9,8	7	0,28	28	14
9,8 – 11,4	4	0,16	16	18
11,4 – 13,0	4	0,16	16	22
13,0 – 14,6	3	0,12	12	25
Итого	25	1,00	100	–

Статистический ряд распределения является обязательным итогом любой группировки.

Стат. ряды распределения могут быть изображены графически. Чаще всего графики используют для изображения вариационных стат. рядов.

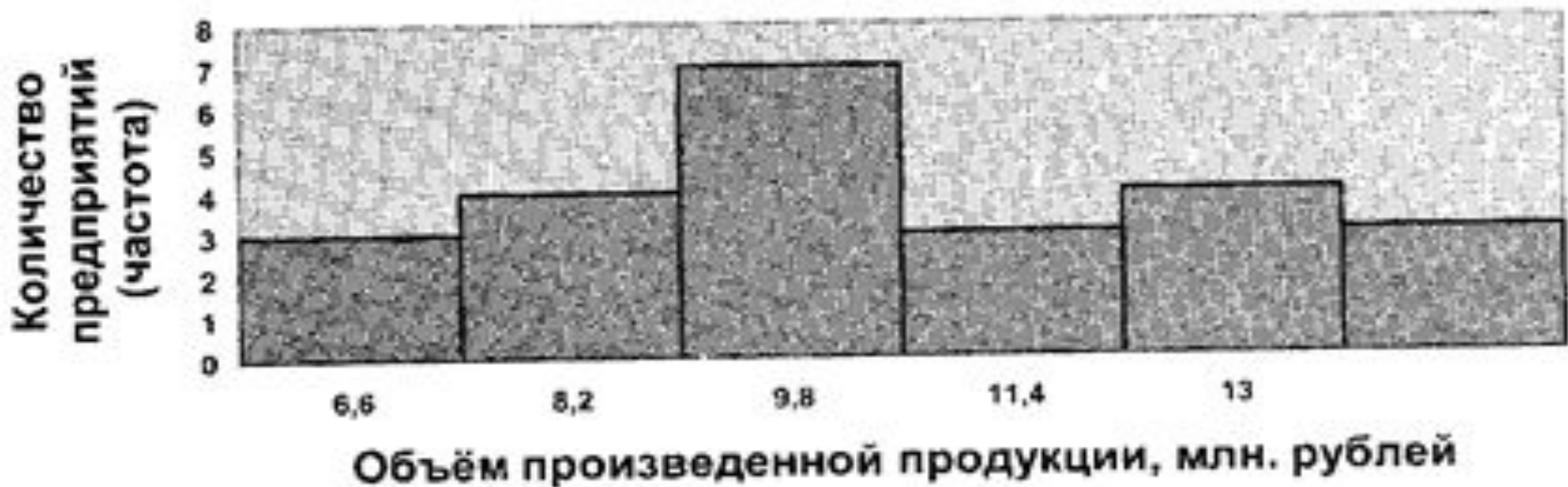
Дискретный ряд изображают с помощью линейной диаграммы, которая называется **полигоном распределения**. При построении его в прямоугольной системе координат по *оси абсцисс* откладываются варианты (значения анализируемого признака), а по *оси ординат* — частота или частость. Целесообразнее по оси ординат откладывать частости: во-первых, это удобнее при большом объеме совокупности, а во-вторых, это позволяет на одном графике изображать статистические ряды распределения нескольких признаков с разным числом единиц совокупности. Полученные на пересечении абсцисс и ординат точки соединяют прямыми линиями, в результате чего получают ломаную линию, называемую полигоном распределения.

На рис. представлен **полигон распределения**, построенный по данным табл. «Распределение рабочих цеха по квалификации»



Интервальный ряд распределения изображается в форме столбиковой диаграммы, которая в этом случае называется **гистограммой**.

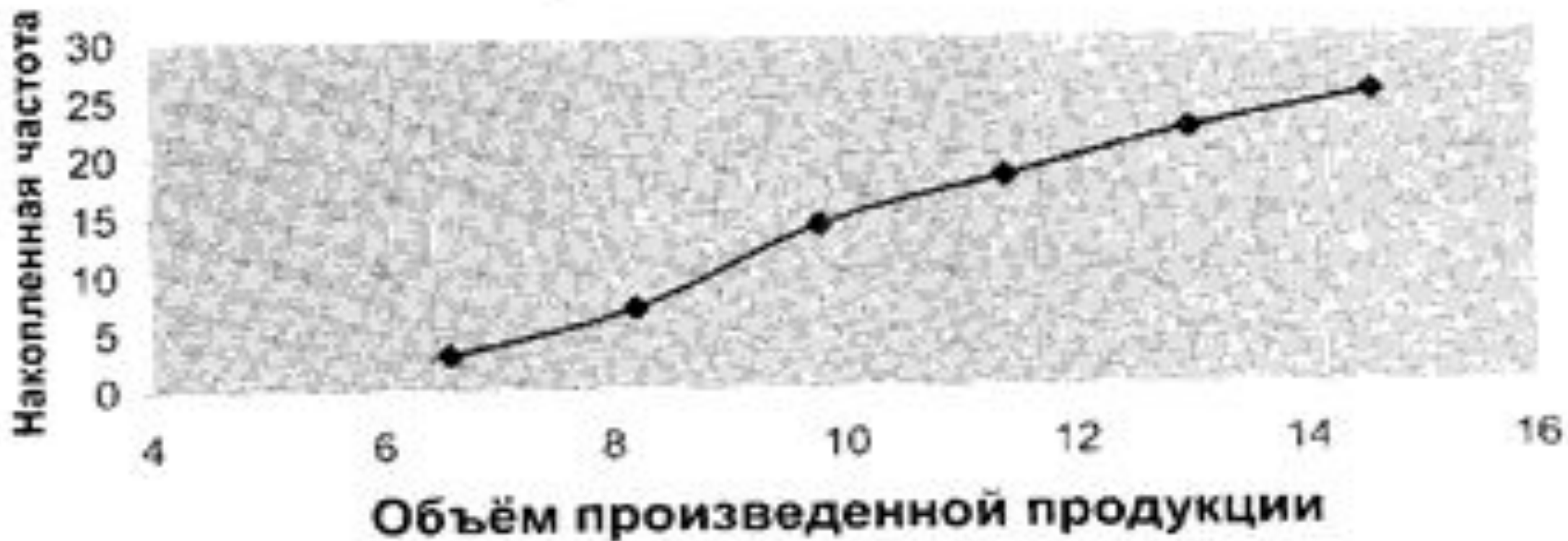
Построим гистограмму по данным табл. «Распределение предприятий по объему произведенной продукции». Для этого по оси абсцисс откладываются равные отрезки, в принятом масштабе соответствующие величине интервалов ряда. На этих отрезках строят прямоугольники, высота которых равна частоте или частости каждого интервала.



Для решения ряда задач (определение структурных средних, наблюдение за процессом концентрации изучаемого явления) ряды распределения преобразовывают в **кумулятивные ряды**, строящиеся по накопленным частотам или частостям.

Накопленные частоты (частости) каждого интервала ряда распределения *рассчитываются как сумма частоты (частости) этого интервала и частот (частостей) всех интервалов, которые ему предшествуют.*

При построении таких рядов в прямоугольной системе координат по оси абсцисс откладываются верхние границы интервалов ряда распределения, а по оси ординат – соответствующие этим интервалам **накопленные частоты (частости)**. Такой ряд, построенный по данным табл. «Распределение предприятий по объему произведенной продукции».



С помощью кумулятивных кривых можно иллюстрировать процесс концентрации, если наряду с накопленными частотами (или частостями) иметь в статистическом ряду распределения суммы накопленных группировочных и других важных признаков изучаемого явления.

Так, в приведенном примере, накопленная частота, например, второго интервала свидетельствует, что 7 предприятий (28% их общей численности), имеющие каждый объем произведенной продукции не более чем 8,2 млн. рублей, производят только 19% всей продукции по совокупности.

Одним из важнейших требований, предъявляемых к статистическим рядам распределения, является *сопоставимость во времени и в пространстве отдельных интервалов*. В рядах с равными интервалами это требование выполняется «автоматически».

Для обеспечения выполнения сопоставимости статистических рядов с неравными интервалами рассчитывается **плотность распределения**, представляющая собой частное от деления частоты интервала на его длину. При графическом изображении таких рядов по оси ординат вместо частот (частостей) откладываются значения плотности распределения.

Существует множество графических изображений. В основу их классификации может быть положен ряд признаков: а) способ построения графического образа; б) геометрические знаки, изображающие статистические показатели и отношения; в) задачи, решаемые с помощью графического изображения.

По способу построения статистические графики делятся на диаграммы и статистические карты. Диаграммы - наиболее распространенный способ графических изображений. *Диаграммы* применяются для наглядного сопоставления в различных аспектах (пространственном, временном и др.) независимых друг от друга величин: территорий, населения и т.д. При этом сравнение исследуемых совокупностей производится по какому-либо существенному варьирующему признаку. *Статистические карты* - графики количественного распределения по поверхности. Они представляют собой условные изображения статистических данных на контурной географической карте, то есть показывают пространственное размещение и пространственную распространенность статистических данных.

Геометрические знаки - это либо точки, либо линии или плоскости, либо геометрические тела. В соответствии с этим, различают графики точечные, линейные, плоскостные и пространственные (объемные).

При построении точечных диаграмм в качестве графических изображений применяются совокупности точек; при построении линейных - применяются линии.

Диаграммы сравнения

Диаграммы сравнения применяются для графического отображения статистических данных с целью их наглядного сопоставления друг с другом в тех или иных разрезах.

Сравнительные диаграммы делятся на:

- а) диаграммы простого сопоставления;
- б) структурные диаграммы;
- в) изобразительные (фигур-знаков).

Диаграммы простого сопоставления дают наглядную сравнительную характеристику статистических совокупностей по какому-либо варьирующему признаку. При этом сопоставляемые совокупности и их части классифицируются по какому-либо атрибутивному или количественному признаку так, что отражаемый диаграммой статистический ряд представляет собой дискретный ряд цифр, на основе которого и строится график.

Диаграммы простого сопоставления делятся на **полосовые** и **столбиковые**. Основной особенностью этих диаграмм является одномерность графического выражения величин варьирующего признака и их одномасштабность для различных столбцов или полос, характеризующих величину отражаемого признака в разных классификационных группах.

На *столбиковых* диаграммах статистические данные изображаются в виде вытянутых по вертикали прямоугольников. Построение столбиковой диаграммы требует применения вертикальной масштабной шкалы. Основания столбиков размещаются на горизонтальной линии, а высота столбиков устанавливается пропорционально изображаемым величинам.

Полосовые диаграммы состоят из прямоугольников, расположенных горизонтально. В этом случае масштабная шкала - *горизонтальная ось*. Принцип их построения тот же, что и в столбиковых.

Для сопоставления изменяющихся во времени показателей, а также при сравнении величин, относящихся к одному и тому же периоду, могут использоваться *квадратные* и *круговые диаграммы*. В отличие от столбиковых или полосовых диаграмм они выражают величину изображаемого явления размером своей площади. Чтобы изобразить квадратную диаграмму, необходимо из сравниваемых статистических величин извлечь квадратные корни, а затем построить квадраты со сторонами, пропорциональными полученным результатам. Круговые диаграммы строятся аналогично. Разница состоит лишь в том, что на графике вычерчиваются круги, радиусы которых пропорциональны квадратному корню из изображаемых величин.

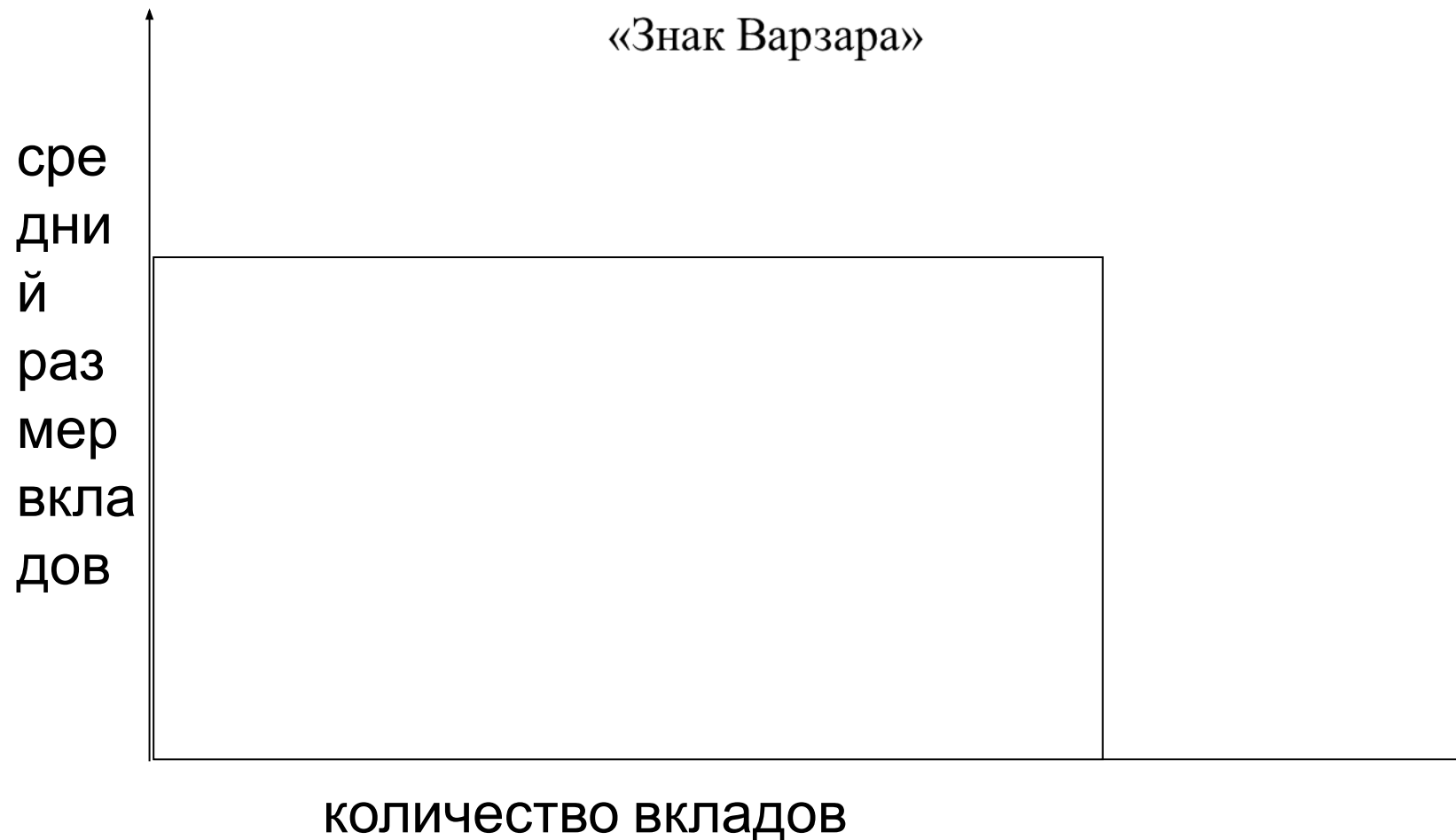
Диаграммы структуры это такие диаграммы, в которых отдельные статистические совокупности сопоставляются по их структуре, характеризующейся соотношением разных параметров совокупности или ее отдельных частей. Простейшим видом структурных статистических диаграмм являются *диаграммы удельных весов*, отражающие структуры сравниваемых совокупностей по процентному соотношению в них отдельных частей, выделяемых по тому или иному количественному или атрибутивному признаку. Эти диаграммы получены путем преобразования простой полосовой диаграммы с подразделенными полосами.

Другой широко распространенный метод графического изображения структур статистических совокупностей по соотношению удельных весов заключается в составлении структурных круговых или секторных диаграмм. *Секторные диаграммы* удобно строить следующим образом: вся величина явления принимается за сто процентов, рассчитываются доли отдельных частей в процентах. Круг разбивается на секторы пропорционально частям изображаемого целого. Таким образом, на 1% приходятся 3,6 градуса. Для получения центральных углов секторов, изображающих доли частей целого, необходимо их процентное выражение умножить на 3,6 градуса.

Секторные диаграммы выглядят убедительно при существенных различиях сравниваемых структур, а при небольших различиях они могут быть недостаточно выразительны. Значительным преимуществом **столбиковых структурных** диаграмм по сравнению с секторными является их большая емкость, возможность отразить на небольшом пространстве большой объем полезной информации.

Для одновременного изображения трех величин, связанных между собой таким образом, что одна величина является произведением двух других, применяются диаграммы, называющиеся *«знаком Варзара»*. «Знак Варзара» представляет собой прямоугольник, у которого один сомножитель принят за основание, другой за высоту, а вся площадь равна произведению.

Оба показателя откладываются на шкалах (каждый на своей), третий (результат) изображается в виде прямоугольника в поле графика.



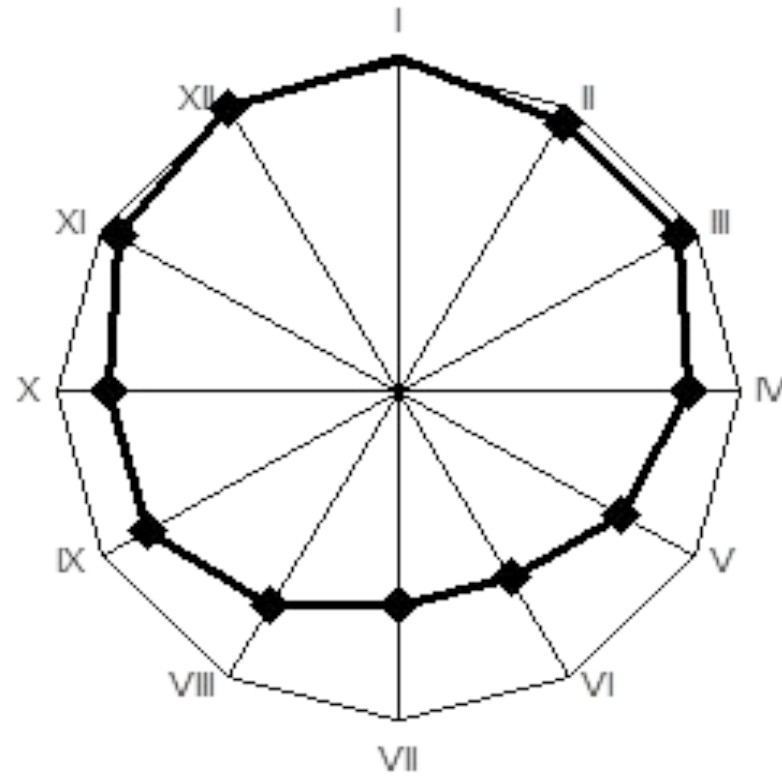
Средний размер вклада, умноженный на их число, дает общую сумму вкладов, что и отображается в виде площади (данные в центре прямоугольников, млрд. руб.).

Для изображения и внесения суждений о развитии явления во времени строятся **диаграммы динамики**. В рядах динамики используются для наглядного изображения явлений многие диаграммы: **столбиковые, ленточные, квадратные, круговые, линейные, радиальные** и другие. Выбор вида диаграмм зависит в основном от особенностей исходных данных, от цели исследования.

Например, если имеется ряд динамики с несколькими неравноотстоящими уровнями во времени (1913, 1940, 1950, 1980, 1985, 1995), то часто для наглядности используют *столбиковые, квадратные или круговые диаграммы*. Они зрительно впечатляют, хорошо запоминаются, но не годны для изображения большого числа уровней, так как громоздки, и если число уровней в ряду динамики велико, то целесообразно применять *линейные диаграммы*, которые воспроизводят непрерывность процесса развития в виде непрерывной ломаной линии. Кроме того, линейные диаграммы удобно использовать: когда целью исследования является изображение общей тенденции и характера развития явления; когда на одном графике необходимо изобразить несколько динамических рядов с целью их сравнения; когда наиболее существенным является сопоставление темпов роста, а не уровней.

К диаграммам динамики относятся и *радиальные диаграммы*, построенные в полярных координатах и предназначенные для отражения процессов, ритмически повторяющихся во времени. Чаще всего эти диаграммы применяются для иллюстрации сезонных колебаний, и в этом отношении они имеют преимущество перед статистическими кривыми.

Радиальная диаграмма



Месячное потребление электроэнергии

**ДЗ! ПРИДУМАТЬ И ОТОБРАЗИТЬ
ГРАФИЧЕСКИ
3 ПРИМЕРА
НА ОСНОВЕ СТАТ.ДАННЫХ**