



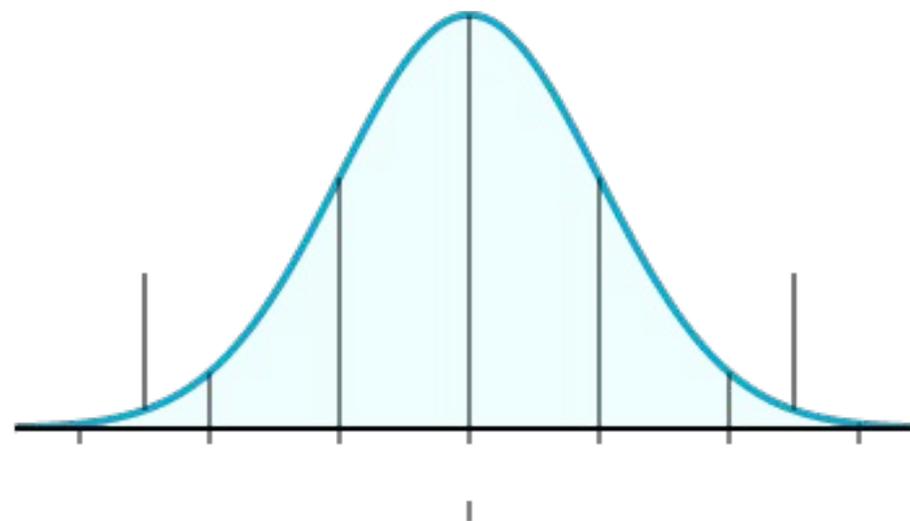
Статистика с Excel & R (+CFA)

Сводка и группировка



Содержание

1. Статистическая сводка
2. Группировка
3. Ряды распределения



1. Статистическая сводка



Статистическая сводка



SCHOOL
OF ECONOMICS
AND FINANCE

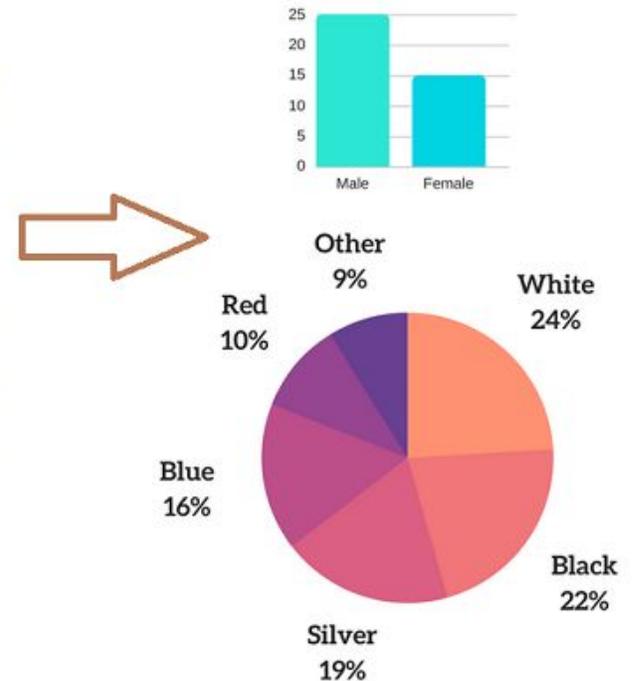
Статистическая сводка (Summary statistics) – метод научной обработки статистических данных, собранных в процессе наблюдения, при котором информация, относящаяся к отдельным единицам, обобщается, а затем характеризуется аналитическими показателями и системой таблиц.

Статистическая сводка

Статистическая сводка – процесс упорядочения, систематизации и обобщения данных статистического наблюдения.

	A	B	C	D
1	Respondent Number	Age	Gender	Favorite Car Color
2	1	22	M	White
3	2	37	F	Silver
4	3	45	F	Black
5	4	62	F	Gray
6	5	28	M	Red
7	6	45	M	Green
8	7	88	F	Brown
9	8	61	M	White
10	9	95	M	Black
11	10	27	M	White
12	11	39	F	Green
13	12	43	M	Brown
14	13	55	F	Black
15	14	59	F	White

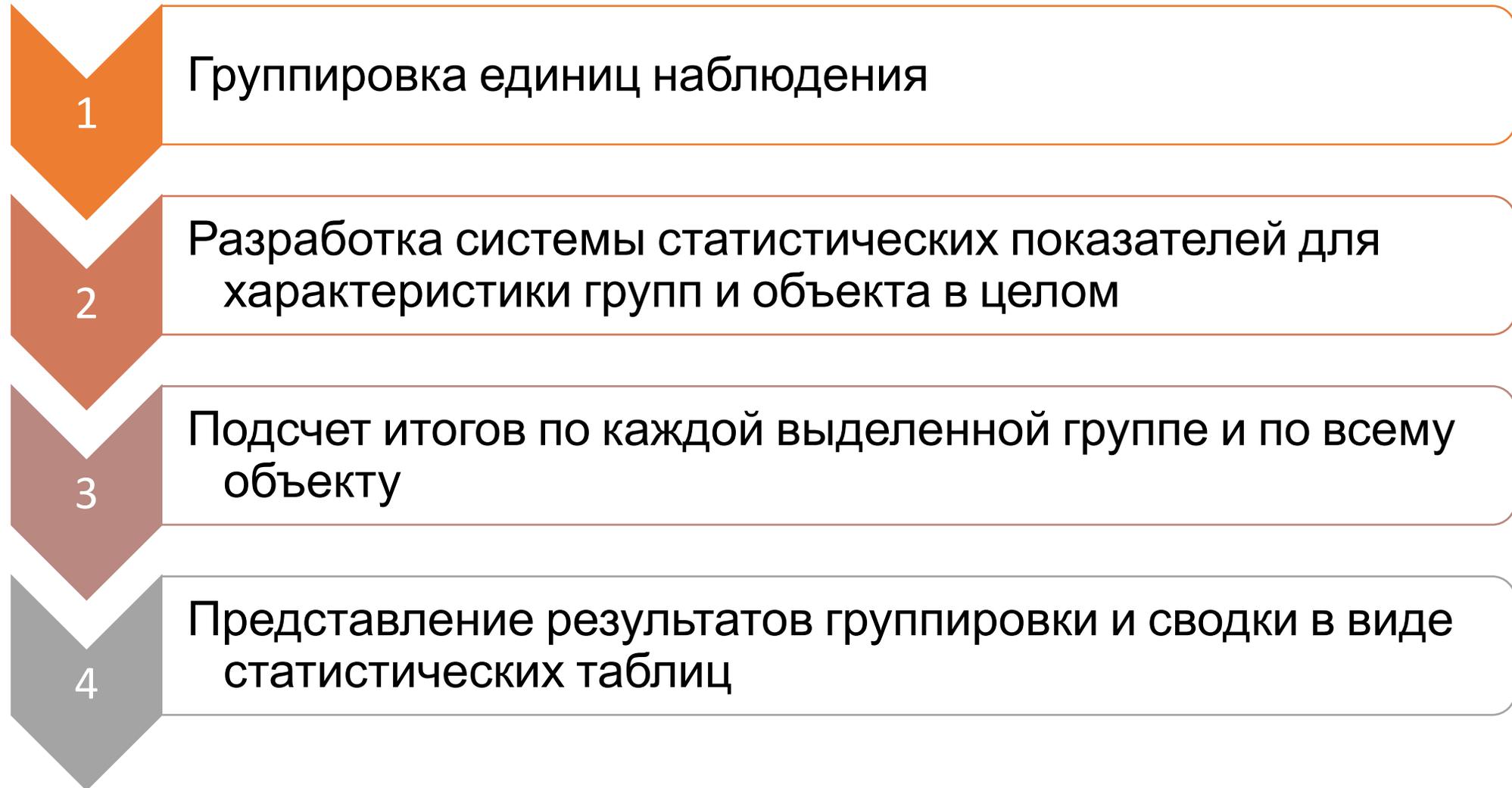
RAW DATA



Descriptive Statistics

Статистическая сводка

Этапы сводки



2. Группировка



Группировка



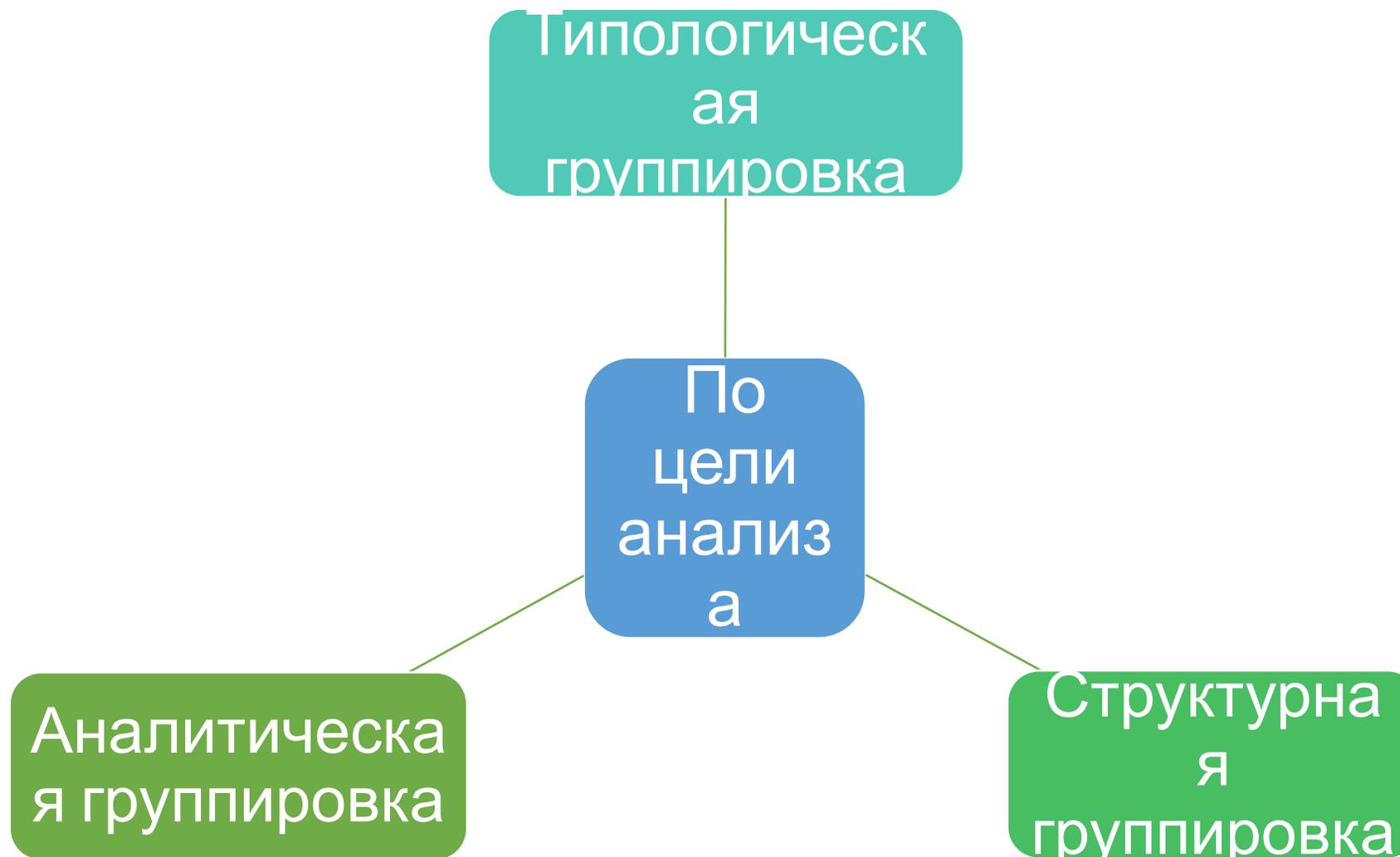
SCHOOL
OF ECONOMICS
AND FINANCE

Группировка (Statistical (Data) grouping) – это метод исследований экономических явлений путем выделения однородных групп по соответствующим определенным признакам, через которые раскрываются существенные черты, особенности состояния и развития всей совокупности.

Групировка



SCHOOL
OF ECONOMICS
AND FINANCE



Группировка



SCHOOL
OF ECONOMICS
AND FINANCE

Типологическая
группировка

Для выявления социально-экономических типов явлений (например, группы населения по роду деятельности, группы предприятий по формам собственности)

По
цели
анализа

Для определения наличия, направления и формы связи между явлениями и процессами.

Аналитическая
группировка

Структурная
группировка

Для изучения структуры социально-экономических явлений и процессов, соотношения их отдельных частей (состав рабочих по профессиям, стажу работы, возрасту, состав себестоимости продукции по видам затрат и т.д.)

Группировка



SCHOOL
OF ECONOMICS
AND FINANCE

По
сложности
построения

Простая

- Изучение взаимосвязи между явлениями, сгруппированными по 1 признаку

Комбини-
рованная
(сложная)

- Изучаемая совокупность делится сначала по 1 признаку, а потом внутри каждой группы по другому признаку и т.д.

Группировка

Группировочный признак (Grouping variable) – это признак, по которому происходит разделение единиц в группе. Его выбор зависит от цели группировки и существа данного явления.

Маленькое количество групп дает грубое представление о характере распределения, интервалы будут слишком большими

Слишком большое количество групп, в силу случайных обстоятельств, тоже не дает правильного представления о характере распределения.



Группировка

Маленькое количество групп дает грубое представление о характере распределения, интервалы будут слишком большими

Слишком большое количество групп, в силу случайных обстоятельств, тоже не дает правильного представления о характере распределения.

Рекомендуется придерживаться следующего порядка при решении вопроса о количестве групп.

При построении группировки по качественному признаку, групп, как правило, будет столько, сколько имеется градаций, видов, состояний у этого признака.

Число предполагаемых групп (n) при количественно выраженных признаках можно определить по формуле Стерджесса:

$$n = 1 + 3,322 \lg N,$$

где N – общая численность ряда распределения (число единиц совокупности).



Группировка



SCHOOL
OF ECONOMICS
AND FINANCE

Интервал (Interval) – это значения варьирующего признака, лежащие в определенных границах. Нижняя граница интервала – наименьшее значение признака в интервале. Верхняя граница – наибольшее значение признака в нем.



Открытые – те, у которых только одна граница, верхняя у первого, нижняя у последнего.



Закрытые – те, у которых обозначены обе границы.

Группировка

Пример 1. Показатели работы акционерных обществ района за отчетный

№ Предприятия	Стоимость основных средств, млн тенге	Выпуск продукции, млн тенге	ГОД		№ Предприятия	Стоимость основных средств, млн тенге	Выпуск продукции, млн тенге
1	7.5	12.4			11	8.6	18.3
2	8.4	12.5			12	8.2	19.6
3	5.7	10.1			13	5.9	9.7
4	8.2	14.3			14	5.1	10.4
5	6.6	12.9			15	5.2	10.3
6	3.4	3.5			16	8.6	20.6
7	1.9	2.5			17	4.7	6.6
8	3.7	4.9			18	3.8	3.6
9	3.8	2.7			19	10.2	18
10	1.2	2.3			20	6.3	9.7

Группировка



SCHOOL
OF ECONOMICS
AND FINANCE

Расчет величины равного интервала производится по формуле:

$$i = \frac{X_{max} - X_{min}}{n}$$

где n – число групп,

X_{max} – наибольшее значение признака в совокупности

X_{min} – наименьшее значение признака в совокупности

$$i = (10.2 - 1.2) / 5 = 1.8$$

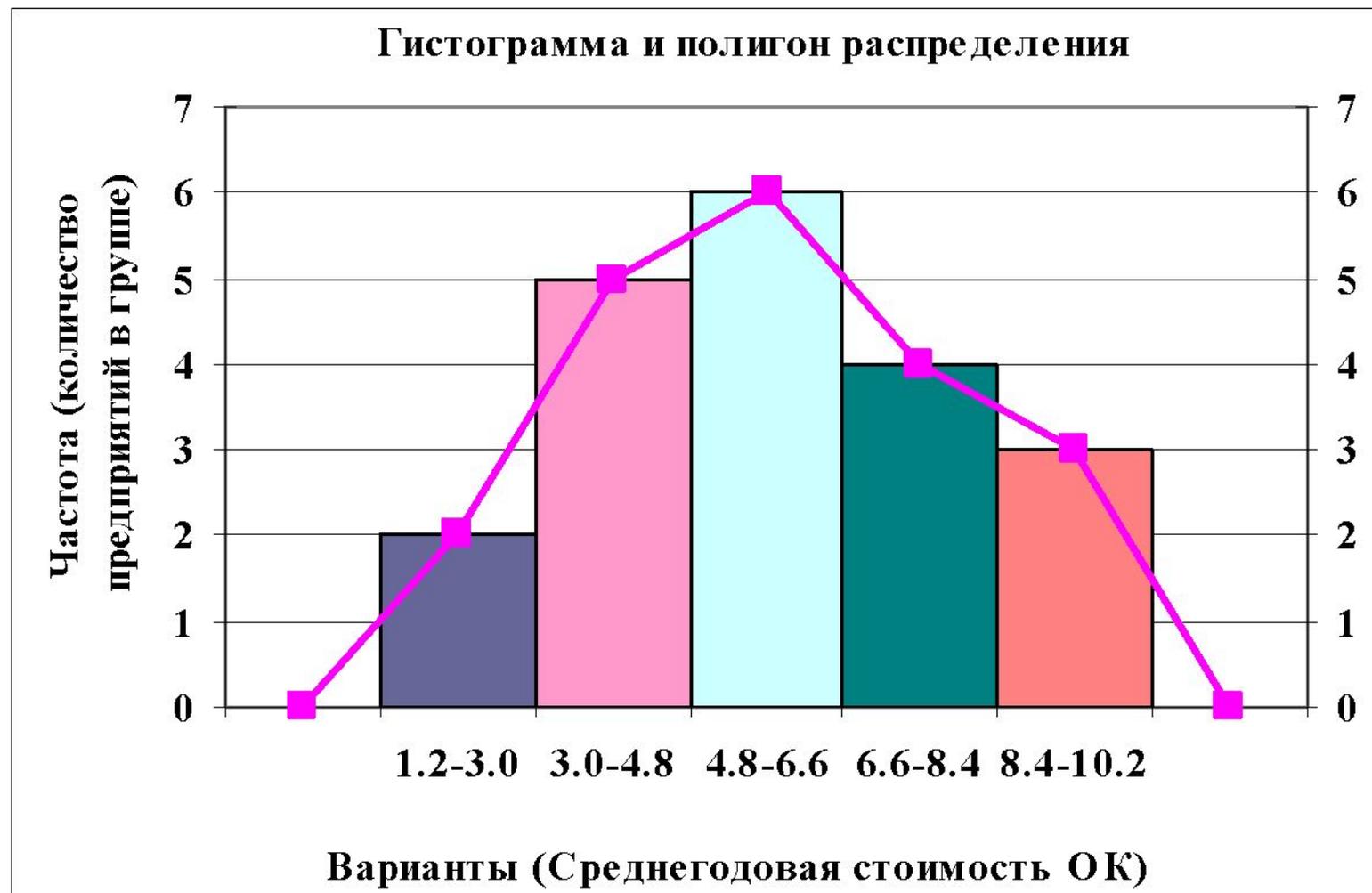
Группировка

Группировка АО по среднегодовой стоимости основного капитала

№ Группы	Интервал	Кол-во предприятий в группе	Стоимость основных средств, млн тенге		Выпуск продукции, млн тенге	
			По группе	В среднем на 1 предприятие	По группе	В среднем на 1 предприятие
1	2	3	4	5=4:3	6	7=6:3
1	1.2 – 3.0	2	3.1	1.55	4.8	2.4
2	3.0 – 4.8	5	19.4	3.88	21.3	4.3
3	4.8 – 6.6	6	34.8	5.80	63.1	10.5
4	6.6 – 8.4	4	32.3	8.08	58.8	14.7
5	8.4 – 10.2	3	27.4	9.13	56.9	19.0
Итого		20	117.0	5.85	204.9	10.2

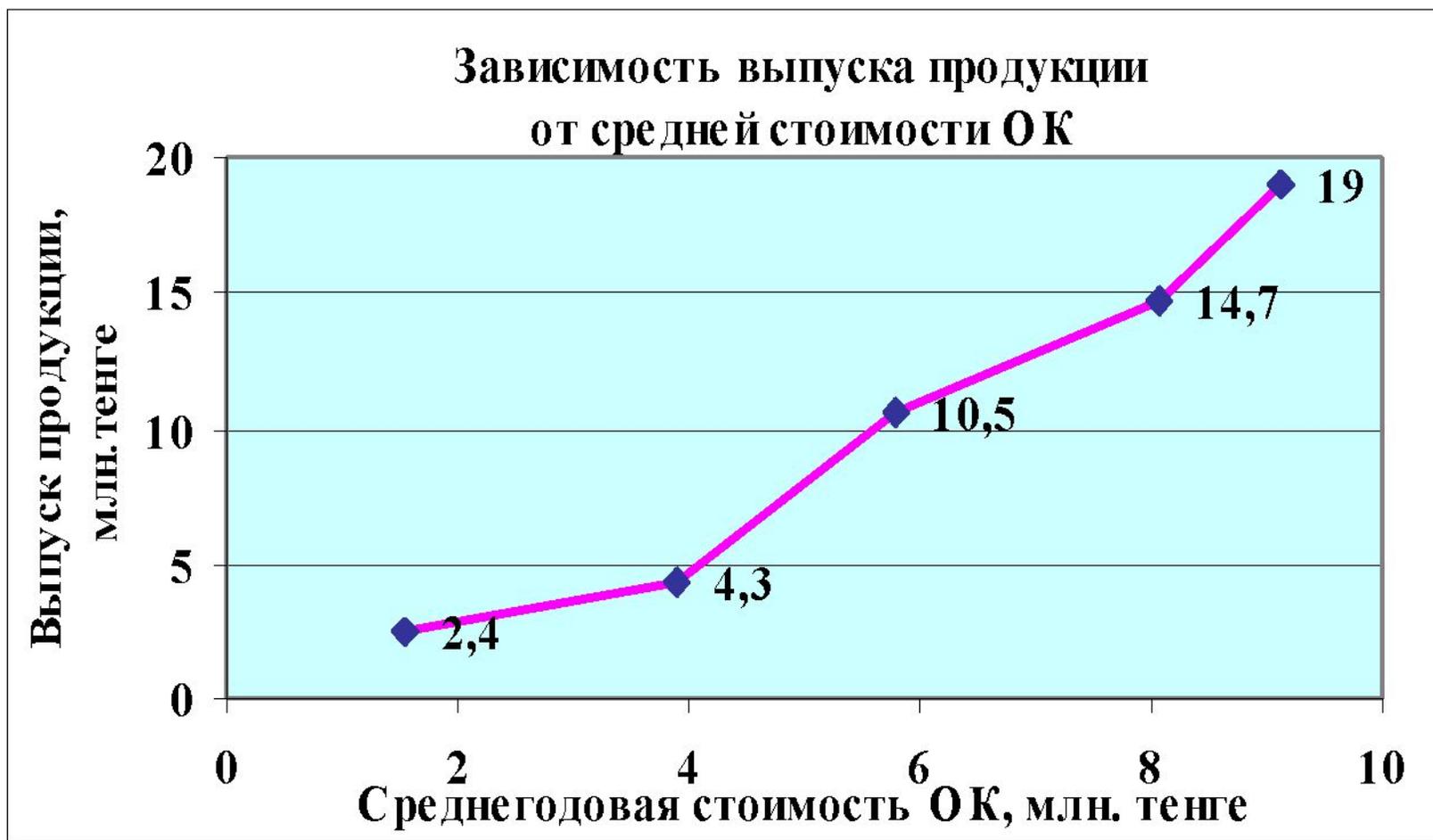
Группировка

Гистограмма и полигон распределения



Группировка

Зависимость выпуска продукции от средней стоимости ОК



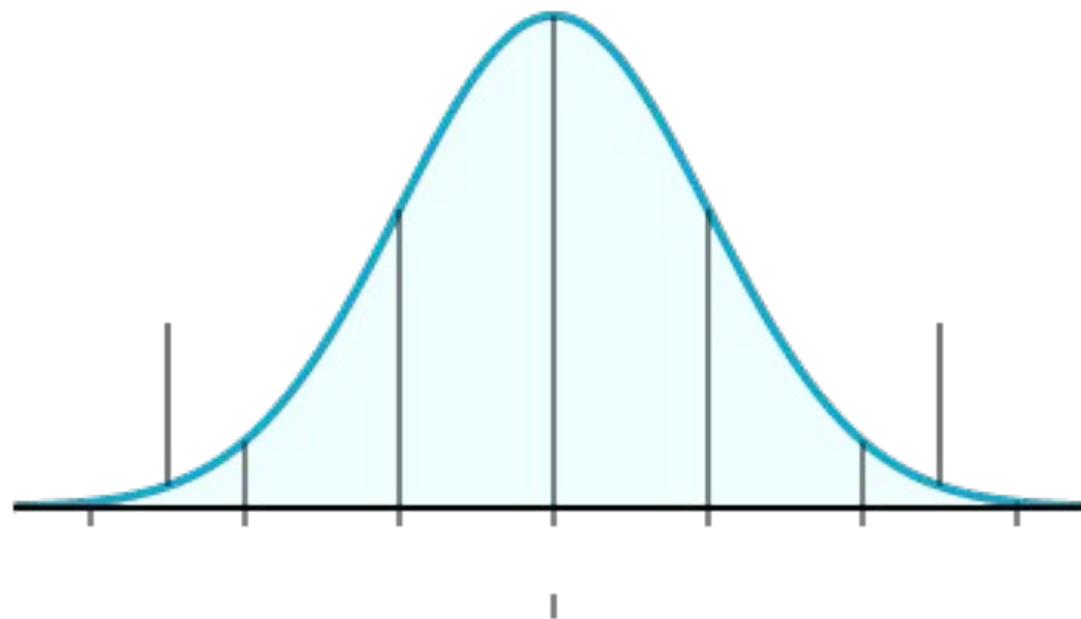
Группировка

ВЫВОД:

Между среднегодовой стоимостью основного капитала и объемом выпуска продукции АО существует прямая положительная зависимость,

Т.е.: с увеличением среднегодовой стоимости ОК выпуск продукции АО увеличивается

3. Ряды распределения



Ряды распределения

Статистический ряд распределения (Distribution) – это упорядоченное распределение единиц совокупности на группы по определенному варьирующему признаку.

Качественный признак

Атрибутивные ряды распределения характеризуют состав совокупности по тем или иным существенным признакам. Взятые за несколько периодов, эти данные позволят исследовать изменение структуры. Атрибутивный ряд распределения принято оформлять в виде таблиц.

Количественный признак

Вариационные ряды распределения состоят из элементов:
вариантов и частот.

Ряды распределения

Вариационные ряды распределения состоят из элементов:
вариантов и частот.

Частоты – частоты, выраженные в долях единицы или в процентах к итогу. Соответственно сумма частостей будет равна 1 или 100%

Варианты – отдельные значения признака, которые он принимает в вариационном ряду, т. е. конкретное значение варьирующего признака (обозначаются x).

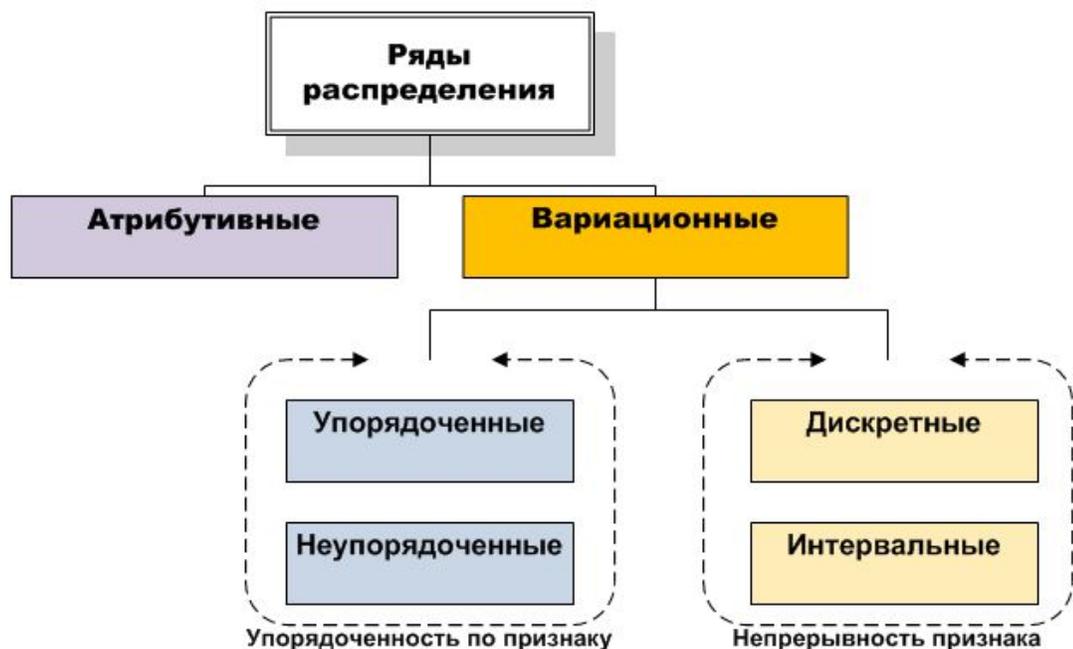
Частоты – это численности отдельных вариантов или каждой группы вариационного ряда, т.е. это числа, показывающие как часто встречаются те или иные варианты в ряду распределения (обозначаются f). Сумма всех частот определяет численность всей совокупности, ее объем.

Ряды распределения

Вариационные ряды распределения

Дискретные – если признак принимает только целые значения.

Интервальные – если значения признака задаются в виде интервалов, в пределах которых они могут принимать любые значения.



Частоты, рассчитанные на единицу величины интервала называются **плотностью распределения признака**.

Ряды распределения

Пример 2. Имеются данные о численности рабочих по тарифным разрядам

Разряд (х - варианты)	Количество рабочих, чел. (f - частоты)	Доля рабочих (%) (f - частоты)
1	4	$8 = (4/50) * 100\%$
2	8	16%
3	11	22%
4	15	30%
5	7	14%
6	5	10%
Итого	50	100%

Ряды распределения

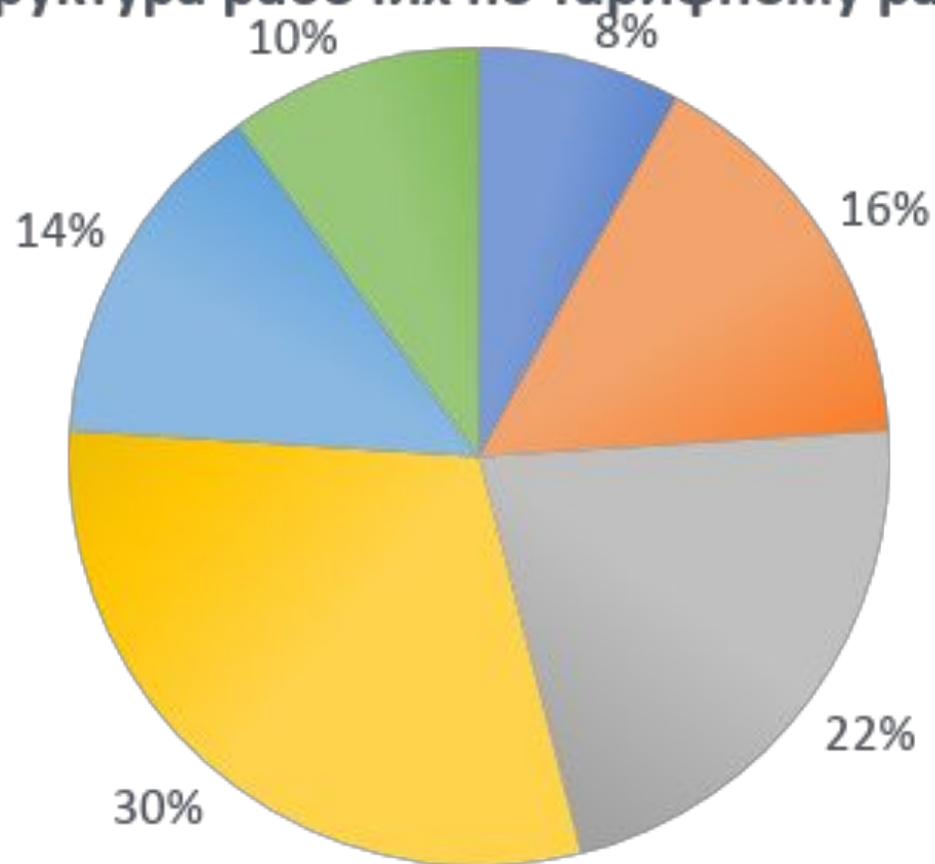
Гистограмма



Ряды распределения

Структура

Структура рабочих по тарифному разряду



- 1
- 2
- 3
- 4
- 5
- 6



ALMA
ALMATY MANAGEMENT
UNIVERSITY



SCHOOL
OF ECONOMICS
AND FINANCE

THANK YOU