

с/л Статистический анализа временных рядов

Преподаватель:

Цеховая Татьяна Вячеславовна,
кандидат физ. – мат. наук, доцент

Лабораторные занятия - **34 часа**

Отчетность - **зачет**

Мероприятия по текущему контролю знаний:

- **лабораторный практикум**

Глава 1.

ОПИСАТЕЛЬНАЯ СТАТИСТИКА

Тема 1.

Ряды распределения.

Одномерные статистические признаки

Вариационные ряды

Опр. Вся подлежащая изучению совокупность объектов называется **генеральной** совокупностью.

Опр. Часть объектов, отобранная для исследования, называется **выборочной** совокупностью (**выборкой**).

Опр. Выборочные данные, полученные в ходе эксперимента, называются **экспериментальными** (**эмпирическими**) данными.

Пример 1. Заданы оценки 30 студентов по специальному курсу «Математические методы анализа данных» в порядке сдачи экзамена:

3 10 4 8 9 8 5 7 10 9 5 7 9 6 5
6 7 6 9 10 8 6 7 6 8 8 6 7 4 8

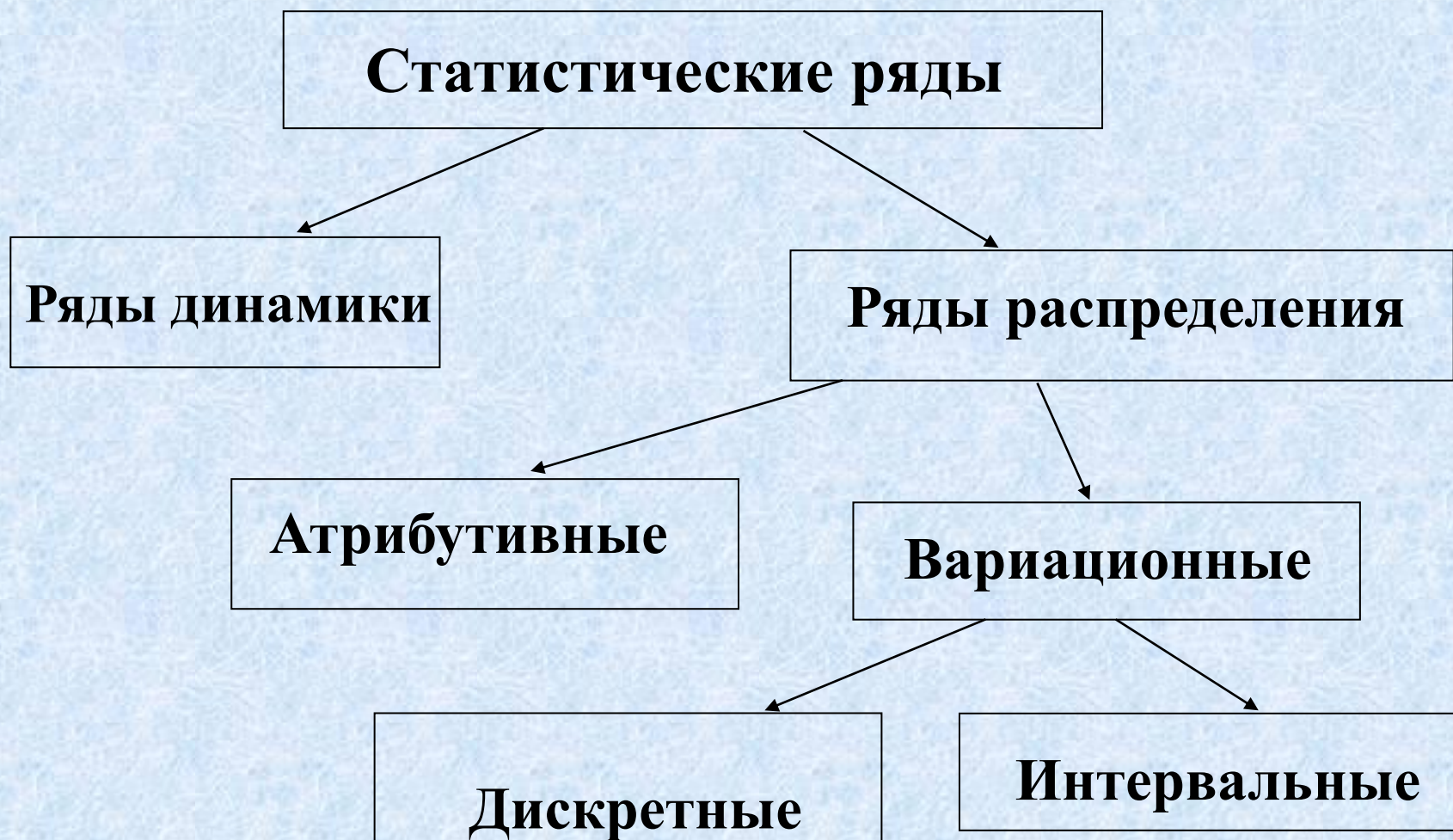
Оценка	3	4	5	6	7	8	9	10
Количество студентов	1	2	3	6	5	6	4	3

Опр. Ряд чисел, характеризующий распределение элементов совокупности по тому или иному признаку, называется **статистическим рядом распределения.**

Статистические ряды распределения можно строить как для количественного признака (**вариационные ряды**), так и для качественного признака (**атрибутивные ряды**).

Если данные систематизированы по времени, то моделью группировки будет **временной ряд** (**ряд динамики**).

КЛАССИФИКАЦИЯ СТАТИСТИЧЕСКИХ РЯДОВ



ПРИМЕР РЯДА ДИНАМИКИ

Выпуск продукции предприятия за 2013 - 2018 годы характеризуются следующими данными (в сопоставимых ценах), млрд. руб.:

Годы	2013	2014	2015	2016	2017	2018
Выпуск продукции	23,3	24,9	26,6	27,6	29,0	32,2

ПРИМЕР АТРИБУТИВНОГО РЯДА РАСПРЕДЕЛЕНИЯ

Распределение работников предприятия по образованию	
Образование работников	Количество работников
высшее	20
неполное высшее	25
среднее специальное	35
среднее	50
Итого	130

Опр. Вариационный ряд – ранжированный в порядке возрастания (убывания) ряд элементов с соответствующими им весами (частотой).

В **дискретном вариационном ряду** (ДВР) частоты распределяются непосредственно по значениям варьирующего признака.

В **интервальном вариационном ряду** (ИВР) частоты распределяются по интервалам группировки.

ПРИМЕР ДИСКРЕТНОГО ВАРИАЦИОННОГО РЯДА

Распределение рабочих по тарифным разрядам:

Тарифный разряд	Число рабочих, чел.
1-й	10
2-й	20
3-й	40
4-й	60
5-й	50
6-й	20
Итого:	200

ПРИМЕР ИНТЕРВАЛЬНОГО ВАРИАЦИОННОГО РЯДА

Распределение нарушений скоростного режима:

Превышение разрешенной скорости движения (км/ч)	20-30	30-40	40-45	45-60	60 и более
Количество нарушений	10	20	15	10	5

АЛГОРИТМ ПОСТРОЕНИЯ ДВР

Рассмотрим одномерный количественный признак. Произведем выборку объема n :

$$X = \{x_1, x_2, \dots, x_n\}.$$

Обозначим:

$$x_{\min} = \min \{x_1, x_2, \dots, x_n\},$$

$$x_{\max} = \max \{x_1, x_2, \dots, x_n\}.$$

Шаг 1.

Определим число k различных значений из числа всех n наблюдений выборки X . Упорядочим эти значения по возрастанию:

$$x(1) < x(2) < \dots < x(k),$$

где $x(i)$ – i -ая **варианта**, $i = 1, \dots, k$.

Очевидно, что

$$x(1) = x_{\min}, \quad x(k) = x_{\max}.$$

Обозначим

$$x = \{ x(1), x(2), \dots, x(k) \}.$$

Шаг 2.

Для каждой варианты $x(i)$ вычислим ее **абсолютную частоту** m_i — число появлений i -ой варианты в выборке X объема n .

Должно выполняться условие нормировки:

$$\sum_{i=1}^k m_i = n.$$

Шаг 3.

Для каждой варианты $x(i)$ вычислим ее **относительную частоту**

$$w_i = \frac{m_i}{n}, \quad \overline{1, k}.$$

Должно выполняться условие нормировки:

$$\sum_{i=1}^k w_i = 1.$$

Дискретный вариационный ряд — это последовательность элементов выборки, записанная в возрастающем порядке, и соответствующих им абсолютных (относительных) частот.

$x(i)$	$x(1)$	$x(2)$	\dots	$x(k)$
m_i	m_1	m_2	\dots	m_k
w_i	w_1	w_2	\dots	w_k

Пример 2. На заводе работает 600 человек. Произвольным образом отобрали 30 человек и выписали их тарифные разряды:

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

Построить ДВР, вычислив абсолютные и относительные частоты этих вариантов.

Решение. $N = 600, \quad n = 30, \quad n \ll N,$

$$X = \{ x_1, x_2, \dots, x_{30} \}.$$

Значениями признака служат числа:

1 2 3 4 5 6

Следовательно,

$$x = \{x(1), x(2), \dots, x(6)\}.$$

Тогда

$x(i)$	1	2	3	4	5	6	Σ
m_i	3	4	5	6	6	6	30
w_i	0,1	0,13	0,17	0,2	0,2	0,2	1

Виды интервалов:

- **равные**, когда разность между максимальным и минимальным значениями в каждом из интервалов одинакова;
- **не равные**, когда разность между максимальным и минимальным значениями в интервалах может быть различна;
- **открытые**, когда имеется только либо верхняя, либо нижняя граница;
- **закрытые**, когда имеются и нижняя, и верхняя границы.

АЛГОРИТМ ПОСТРОЕНИЯ ИВР с равными интервалами

Шаг 1.

Определяются

$$x_{min} = \min_{i=1,n} x_i,$$

$$x_{max} = \max_{i=1,n} x_i$$

Вычисляется **размах** (**диапазон**) выборки:

$$R = x_{max} - x_{min}.$$

Шаг 2.

Диапазон значений признака $[x_{min}, x_{max}]$ разбивается на k интервалов одинаковой длины.

Число интервалов k определяется одним из следующих способов:

а) **субъективно** (задается непосредственно исследователем);

б) **по формуле Стерджесса:**

$$k = 1 + [3,322 \lg n] = 1 + [\log_2 n],$$

где n – объем выборки, $[\cdot]$ – целая часть числа.

с) при большом объеме выборки задается исследователем с помощью таблицы:

Объем выборки, <i>n</i>	25-40	40-60	60-100	100-200	200 и более
Число интервалов, <i>k</i>	5-6	6-8	7-10	8-12	10-15

д) по умолчанию во многих ППП по статистическому анализу данных: $6 < k < 20$, k – округленное значение $5 \lg n$.

Шаг 3.

Определяется величина интервалов $h = R / k$.

Шаг 4.

Вычисляются границы интервалов (a_i, a_{i+1}) .

1-й способ.

$$a_1 = x_{min}, \quad a_2 = a_1 + h, \dots,$$

$$a_{k+1} = a_k + h = a_1 + kh.$$

2-й способ.

$$a_1 = x_{min} - h/2, \quad a_2 = a_1 + h, \dots, \quad a_{k+1} = a_1 + kh.$$

Шаг 5.

Подсчитывается абсолютная частота m_i – число вариантов, попавших в i - ый интервал (a_i, a_{i+1}) .

(a_i, a_{i+1})	$[a_1, a_2)$	$[a_2, a_3)$	\dots	$[a_{k-1}, a_k)$	$[a_k, a_{k+1}]$
m_i	m_1	m_2	\dots	m_{k-1}	m_k
w_i	w_1	w_2	\dots	w_{k-1}	w_k

Замечания.

- Если крайние интервалы не замкнуты (например, имеют вид “менее 10”, “более 60”), то ширина первого интервала приравнивается к ширине второго, а ширина последнего — к ширине предпоследнего.
- При построении ИВР с неравными интервалами используется понятие **плотности частот**.

Пример 3. Известна выработка (в %) каждого рабочего по отношению к предыдущему месяцу. Всего 30 рабочих.

125	91	82	93	101	111	109	103	102	80
79	105	115	95	84	130	104	117	127	107
85	75	96	104	126	113	98	86	113	123

Построить ИВР. Для определения числа групп воспользоваться формулой Стерджесса.

Решение.

1. Упорядочим ряд:

75	79	80	82	84	85	86	91	93	95
96	98	101	102	103	104	104	105	107	109
111	113	113	115	117	123	125	126	127	130

Найдем

$$x_{\max} = 130, \quad x_{\min} = 75, \quad R = 130 - 75 = 55.$$

2. Вычислим количество групп:

$$k = 1 + [\log_2 30] = 5.$$

3. Определим $h = R / k = 55 / 5 = 11$.

Замечание. Если исходные данные ряда имеют знак после запятой, то h округляют до этого знака.

4. Вычислим границы интервалов:

[75; 86), [86; 97), [97; 108), [108; 119), [119; 130]

5. Подсчитаем абсолютные частоты и составим таблицу:

(a_i, a_{i+1})	[75; 86)	[86; 97)	[97; 108)	[108; 119)	[119; 130]	Σ
m_i	6	5	8	6	5	30
w_i	6/30	5/30	8/30	6/30	5/30	1