

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ  
ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ им. А. И. ГЕРЦЕНА**

Институт информационных технологий и технологического образования  
Кафедра компьютерные технологии и электронного обучения  
Основная профессиональная образовательная программа  
Направление подготовки 09.03.01 Информатика и вычислительная техника  
Направленность (профиль) «Технологии разработки программного  
обеспечения»  
форма обучения - очная

**ЛАБОРАТОРНАЯ РАБОТА №3  
по дисциплине: «Анализ данных и основы Data science»  
ВАРИАЦИОННЫЕ РЯДЫ**

Руководитель:  
кандидат педагогических наук, доцент,  
Светлана Викторовна Гончарова

Автор работы студент 2 курса  
1 группы 1 подгруппы  
Чирцов Тимофей Александрович

Санкт-Петербург  
2022

Цель: вычислить математические характеристики вариационного ряда

Оборудование: Персональный компьютер, Excel

### Задание 1

(3) Постановка задачи:

**Задание № 1.**  
В качестве изучаемого признака рассматривается число продаж каждого из 26 случайно выбранных продавцов универсама:

16, 12, 15, 15, 23, 9, 15, 13, 14, 14, 21, 15, 14, 17, 27, 15, 16, 12, 16, 19, 14, 16, 17, 13, 14, 14.

Необходимо сделать следующее:

- 1) Построит вариационный ряд;
- 2) Провести анализ построенного вариационного ряда:

(4) Результат выполненной работы:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	16	12	15	15	23	9	15	13	14	14	21	15	14
2	17	27	15	16	12	16	19	14	16	17	13	14	14
3													
4	xi	9	12	13	14	15	16	17	19	21	23	27	
5	mi	1	2	2	6	5	4	2	1	1	1	1	
6	wi	0,03846	0,08	0,0869565	0,28571	0,33333	0,4	0,33333	0,25	0,33333	0,5	1	
7													
8	n	26											
9	k	5,56134	6										
10													
11	xmin	9											
12	xmax	27											
13	Δ	3											

### Задание 2

(3) Постановка задачи:

**Задание № 2.**  
Менеджер большого универсама зафиксировал суммы денег, которые израсходовали 184 покупателя, посетившие отдел верхней одежды в день сезонной распродажи по сниженным ценам. Зная минимальную и максимальную стоимости покупки, менеджер сгруппировал данные о суммах, израсходованных на покупки, в виде таблицы (таблица 1).

Таблица 1.

Суммы денег, израсходованные на покупки товаров в отделе верхней одежды (у.ед.)

Интервалы расходов	100-300	300-500	500-700	700-900	900-1100	1100-1300
Число покупателей (mi)	30	38	50	31	22	13
Доля покупателей (wi)	0.163	0.207	0.272	0.168	0.120	0.070

	A	B	C	D	E	F	G	H	I	J	K
1	Интервалы расходов	100-300	300-500	500-700	700-900	900-1100	1100-1300			n	184
2	Число покупателей (mi)	30	38	50	31	22	13				
3	Доля покупателей (wi)	0,163	0,207	0,272	0,168	0,12	0,07				

### Вопрос 1

Какова нижняя и верхняя границы интервалов (по таблице 1)?

Нижняя граница	100	300	500	700	900	1100	1 задание
Верхняя граница	300	500	700	900	1100	1300	

### Вопрос 2

Изучите распределение, представленное в таблице 2. Определить начало первого интервала и правую границу последнего интервала.

Интервалы	Число регионов			
До 60	10			
60-70	29			
70-80	2			
80-90	13			
90-100	-			
Свыше 100	6			
Анализируя данную таблицу, можно сделать вывод, что интервал необходимо брать, как у известных значений, то есть в данном случае этот интервал равен 10. Значит начало первого интервала будет с 50 до 60. А значение последнего доходит до 110				

### Вопрос 3

Для данных Задания №1 определите оптимальную величину интервала и представьте ряд из этого задания в виде интервального ряда.

15	Интервалы	9-12	12-15	15-18	18-21	21-24	24-27	3 задание
16	Частоты xi	1	10	11	1	2	1	
17	Доля xi	0,0385	0,3846	0,4231	0,0385	0,0769	0,0385	

### Вопрос 4

Интервальные ряды бывают с равными и неравными интервалами. Иногда при группировке с равными интервалами сначала определяют число интервалов (групп)  $z$  при заданном объеме совокупности, используя формулу:

$$L = 2\ln(n),$$

И тогда  $k$  в формуле Стеджерса вычисляется по формуле

$$k = \frac{x_{max} - x_{min}}{L}$$

Для Задания 1 вычислите оптимальную величину интервала по данной формуле и сравните его с интервалом, вычисленным по формуле Стеджерса.

L	6,51619	7	4 задание
K	2,57143		

### Вопрос 5

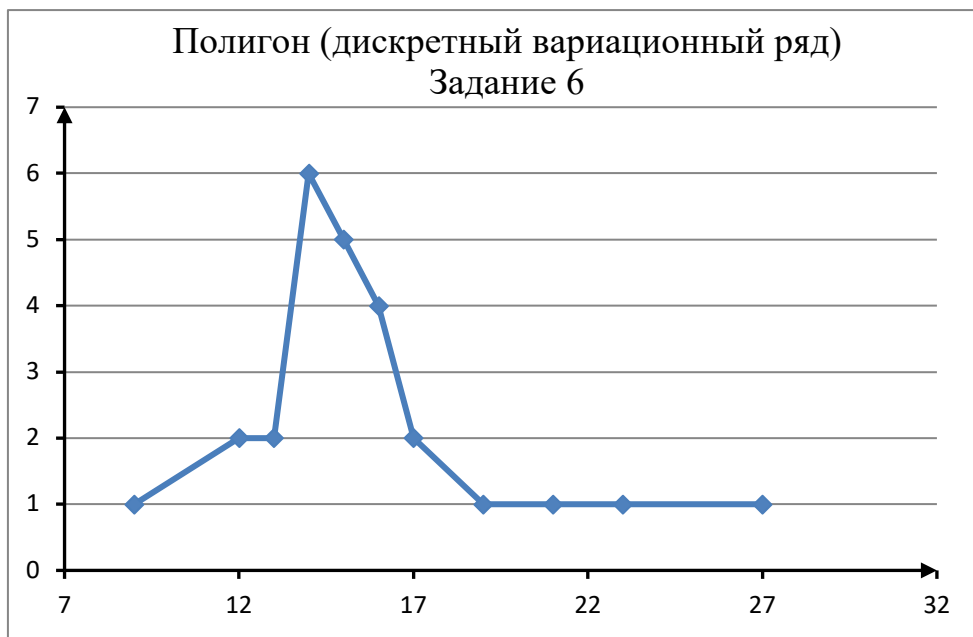
Для данных таблицы 1 вычислите накопленные частоты и расположите их в таблице в восходящем и в нисходящем порядке. На что они указывают? Поясните.

xi	100	300	500	700	900	1100	1300										
Восходящий порядок	0	30	68	118	149	171	184	5 задание	Накопление частоты в восходящем порядке могут указывать на количество значений, которые меньше левого края соответствующего интервала.								
Нисходящий порядок	184	171	149	118	68	30	0		Накопление частоты в нисходящем порядке могут указывать на количество значений, которые больше левого края соответствующего интервала.								

### Вопрос 6

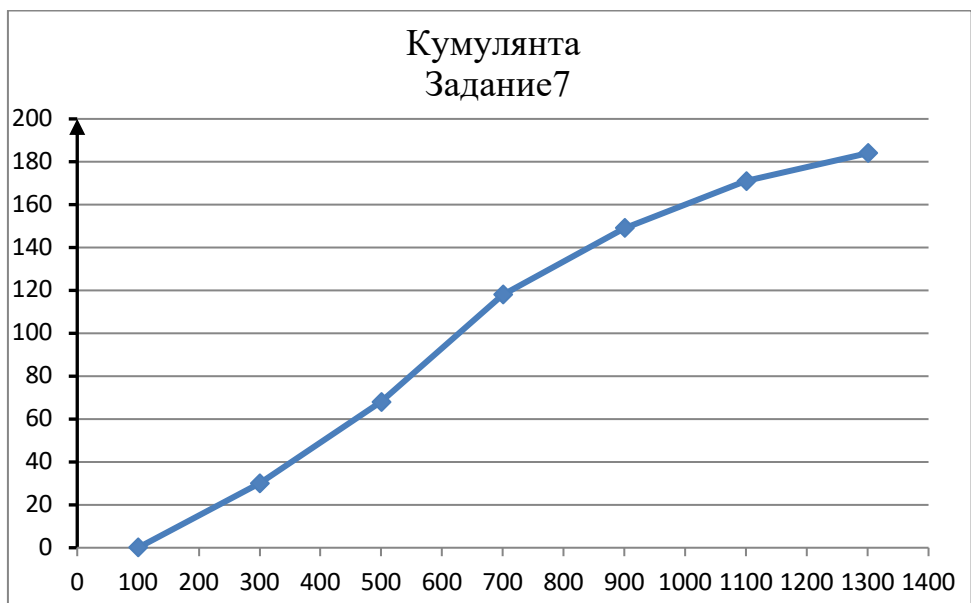
Постройте полигон распределения для Задания 1

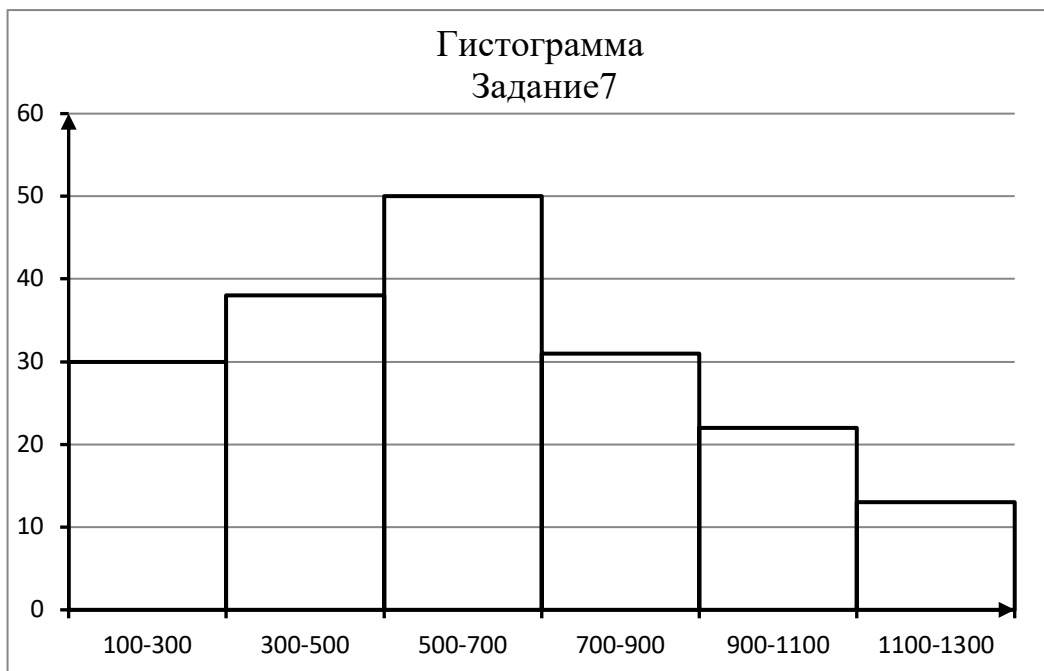




### Вопрос 7

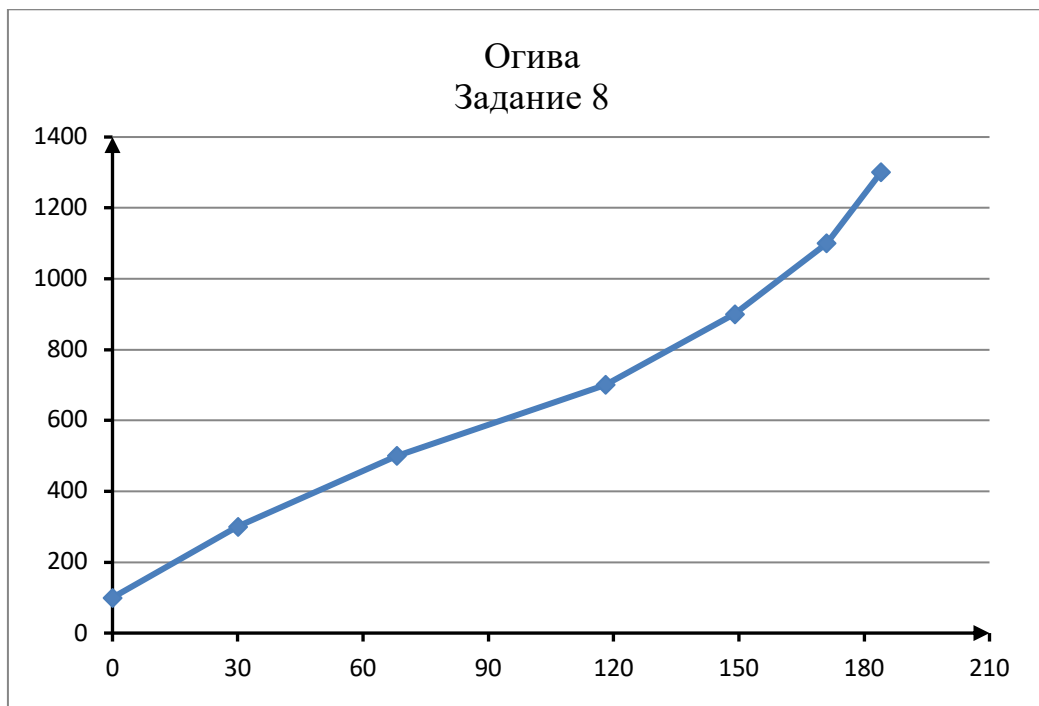
Постройте гистограмму и кумулянту для данных таблицы 2





### Вопрос 8

Постройте огиву для данных таблицы 2. Огива строится аналогично кумулянте (но по оси абсцисс откладываются накопленные частоты (частности), а по оси ординат – значения признака).



### Вопрос 9

Для данных Задания №1 вычислить:

25-й, 50-й, 90-й перцентили в вариационном ряду

9	12	12	13	13	14	14	14	14	14	14	15	15							
15	15	15	16	16	16	16	17	17	19	21	23	27							
25 перцентиль	6,75	Эта позиция находится между 6 и 7 вариантами. Значит мы видим, что значение 6-го по порядку варианта равно 14. Значение 7-го по порядку варианта также равно 14. Отсюда 25-й перцентиль равен 14.																	
50 перцентиль	13,5	Эта позиция находится между 13 и 14 вариантами. Значит мы видим, что значение 13-го по порядку варианта равно 15. Значение 14-го по порядку варианта также равно 15. Отсюда 50-й перцентиль равен 15.																	
90 перцентиль	24,3	Эта позиция находится между 24 и 25 вариантами. Значит мы видим, что значение 24-го по порядку варианта равно 21. Значение 25-го по порядку варианта равно 23. Отсюда 90-й перцентиль - это $21 + (23 - 21) * 0,3 = 21,6$ (0,3 - это расстояние между 24 и 25 значениями)																	

9  
задание

## Вопрос 10

По данным таблицы 2 вычислите медиану (используйте формулу для нахождения медианы внутри медианного интервала)

$$Me = x_{Me(min)} + k_i \frac{0,5 \cdot \sum m_i - v_{Me-1}}{m_{Me}} \quad (2),$$

где  $x_{Me(min)}$  - нижняя граница медианного интервала;

$k_i$  - величина медианного интервала (интервальная разность);

$v_{Me-1}$  - накопленная частота или частость интервал, предшествующего медианному;

$0,5 \sum m_i$  - половина суммы всех частот (или частостей);

$m_{Me}$  - частота медианного интервала.

## Вопрос 11

По данным таблицы 2 вычислить моду (используйте формулу для нахождения моды внутри модального интервала)

$$Mo = x_{Mo(min)} + k_i \frac{m_{Mo} - m_{Mo-1}}{(m_{Mo} - m_{Mo-1}) + (m_{Mo} - m_{Mo+1})}, \quad (3)$$

где  $x_{Mo(min)}$  - нижняя граница модального интервала;

$m_{Mo}$  - частота модального интервала;

$m_{Mo-1}$  - частота интервала, предшествующего модальному;

$m_{Mo+1}$  - частота интервала, последующего за модальным;

$k_i$  - величина модального интервала.

Me	596	10 задание
Mo	577,4194	11 задание

## Вопрос 12

Вычислите среднюю арифметическую для данных Задания №1 по формулам:

- ☐ Средней арифметической
- ☐ Средней арифметической взвешанной

Среднее арифметическое	15,6154	12 задание
Среднее арифметическое взвешанное	16,0385	

### Вопрос 13

Проверить насколько медиана и средняя арифметическая чувствительна к положению крайних значений ряда (для данных таблицы Задания №1)

Так как мода необходимо выбирать по наибольшему значению с графика дискретный вариационный ряд, то	Mo	14
Медиана - это среднее значение ряда, значит, значение медианы будет равно 15, так как это число находится в середине ряда	Me	15
Медиана не зависит от значений признака на краях ранжированного ряда. Медиана зависит от количества значений, а не от того, что показывает само значение. Поэтому часто медиану используют как более надежный показатель типичного значения признака	Средняя арифметическая зависит от количества значений в ряду. Если количество значений будет изменяться, то соответственно и будет изменяться среднее арифметическое значение ряда	

## Вопрос 14

Может ли быть в одном ряду несколько мод? Обоснуйте.

10	10	14	20	10	15	2	14	15	15	3	20	45	Так как мода представляет собой наиболее часто встречающееся значение (в данном задании это значения 10 и 15 - они встречаются пять раз), значит в одном ряду может быть	14 заданий
12	18	20	10	18	10	15	14	18	21	23	20	15		
xi	2	3	10	12	14	15	18	20	21	23	45			
mi	1	1	5	1	3	5	3	4	1	1	1			

Вывод по всей лабораторной работе: С помощью электронных таблиц мы реализовали вычисления математических характеристик вариационного ряда. Ответили на ряд вопросов и получили некоторые выводы по заданным вопросам.