

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МО ЭВМ

ОТЧЕТ
по лабораторной работе №1
по дисциплине «СТАТИСТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ
ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ»
Тема: Формирование и первичная обработка выборки.
Ранжированный и интервальный ряды

Студент гр. 5381

Лянгузов А. А.

Преподаватель

Середа В.И.

Санкт-Петербург

2019

Цель работы.

Ознакомление с основными правилами формирования выборки и подготовки выборочных данных к статистическому анализу.

Задание.

Осуществить формирование репрезентативной выборки заданного объема из имеющейся генеральной совокупности экспериментальных данных.

Осуществить последовательное преобразование полученной выборки в ранжированный, вариационный и интервальный ряды. Применительно к интервальному ряду построить и отобразить графически полигон, гистограмму и эмпирическую функцию распределения для абсолютных и относительных частот. Полученные результаты содержательно проинтерпретировать.

Основные теоретические положения.

Ранжированный ряд - это распределение отдельных единиц совокупности в порядке возрастания или убывания исследуемого признака. Ранжирование позволяет легко разделить количественные данные по группам, сразу обнаружить наименьшее и наибольшее значения признака, выделить значения, которые чаще всего повторяются.

Вариационный ряд - последовательность значений заданной выборки $x^m = (x_1, \dots, x_m)$, расположенных в порядке неубывания:

$$x^{(1)} \leq x^{(2)} \leq \dots \leq x^{(m)}.$$

Интервальный ряд распределения – это таблица, состоящая из двух столбцов (строк) – интервалов варьирующего признака X_i и числа единиц совокупности, попадающих в данный интервал (частот - f_i), или долей этого числа в общей численности совокупностей (частостей - d_i).

Полигоном частот называют ломанную, отрезки которой соединяют точки $(x_1; n_1), (x_2; n_2), \dots, (x_k; n_k)$. Для построения полигона частот на оси абсцисс

откладывают варианты x_i , а на оси ординат - соответствующие им частоты n_i . Точки $(x_i; n_i)$ соединяют отрезками прямых и получают полигон частот.

Гистограммой частот называется ступенчатая фигура, состоящая из прямоугольников с основаниями, равными интервалам значений h_i , и высотами, равными отношению частот к шагу:

$$\frac{m_i}{h_i} \left(\frac{\omega_i}{h_i} = \frac{m_i}{n \cdot h_i} \right).$$

Эмпирической функцией распределения, построенной по выборке $x^m = (x_1, \dots, x_m)$ объема m , называется случайная функция $\hat{F}_m(x)$ равная

$$\hat{F}_m(x) = \frac{1}{m} \sum_{i=1}^m I_{\{x_i \leq x\}}.$$

Значения эмпирической функции распределения принадлежат отрезку $[0,1]$.

Экспериментальные результаты.

Результаты формирования репрезентативной выборки заданного объема из имеющейся генеральной совокупности экспериментальных данных представлены в таблице 1.

Объём выборки: 107.

Номер первого элемента генеральной совокупности экспериментальных данных: 1.

v	501.00	369.00	344.00	473.00	426.00	528.00	497.00	467.00	506.00	431.00	454.00
E	130.40	84.30	86.80	137.90	121.10	163.40	147.30	140.50	158.40	125.00	131.10
v	371.00	482.00	393.00	441.00	463.00	440.00	481.00	340.00	468.00	397.00	496.00
E	89.20	139.90	103.20	122.80	129.10	128.50	135.20	85.10	142.00	108.60	143.10
v	434.00	541.00	352.00	438.00	453.00	423.00	351.00	525.00	409.00	469.00	386.00
E	122.30	146.80	87.70	134.90	119.50	131.10	89.00	165.90	121.00	131.50	95.50

v	505.00	436.00	488.00	449.00	493.00	512.00	472.00	423.00	465.00	351.00	359.00
E	137.50	114.30	134.10	124.50	129.70	169.90	134.20	130.80	140.70	102.90	71.90
v	457.00	467.00	400.00	418.00	492.00	434.00	510.00	392.00	463.00	459.00	397.00
E	126.40	135.10	114.60	118.60	137.50	110.50	140.60	82.70	125.00	145.40	106.80
v	424.00	436.00	429.00	398.00	493.00	522.00	518.00	463.00	437.00	386.00	493.00
E	119.00	116.70	112.90	109.00	154.50	154.50	144.40	121.20	121.80	105.80	151.20
v	414.00	480.00	585.00	562.00	508.00	421.00	463.00	422.00	406.00	544.00	345.00
E	113.50	153.90	177.70	175.90	159.00	117.80	136.70	122.90	110.10	166.70	95.90
v	478.00	393.00	437.00	448.00	458.00	422.00	468.00	430.00	371.00	543.00	471.00
E	126.60	122.80	115.10	121.90	121.70	115.70	144.90	104.30	91.90	155.40	143.90
v	475.00	521.00	353.00	437.00	362.00	490.00	484.00	459.00	480.00	482.00	522.00
E	132.00	139.60	98.00	118.40	111.70	139.90	140.40	136.70	153.30	148.20	143.80
v	576.00	390.00	514.00	442.00	421.00	443.00	438.00	429.00			
E	166.40	91.40	153.60	115.40	107.90	121.90	126.70	120.90			

Таблица 1.

Обработка результатов эксперимента.

Преобразование полученной выборки в ранжированный ряд представлено в таблице 2-3.

340	344	345	351	351	352	353	359	362	369
371	371	386	386	390	392	393	393	397	397
398	400	406	409	414	418	421	421	422	422
423	423	424	426	429	429	430	431	434	434
436	436	437	437	437	438	438	440	441	442
443	448	449	453	454	457	458	459	459	463
463	463	463	465	467	467	468	468	469	471
472	473	475	478	480	480	481	482	482	484
488	490	492	493	493	493	496	497	501	505
506	508	510	512	514	518	521	522	522	525
528	541	543	544	562	576	585			

Таблица 2. Ранжированный ряд для величины y .

71.9	82.7	84.3	85.1	86.8	87.7	89.0	89.2	91.4	91.9
95.5	95.9	98.0	102.9	103.2	104.3	105.8	106.8	107.9	108.6
109.0	110.1	110.5	111.7	112.9	113.5	114.3	114.6	115.1	115.4
115.7	116.7	117.8	118.4	118.6	119.0	119.5	120.9	121.0	121.1
121.2	121.7	121.8	121.9	121.9	122.3	122.8	122.8	122.9	124.5
125.0	125.0	126.4	126.6	126.7	128.5	129.1	129.7	130.4	130.8
131.1	131.1	131.5	132.0	134.1	134.2	134.9	135.1	135.2	136.7
136.7	137.5	137.5	137.9	139.6	139.9	139.9	140.4	140.5	140.6
140.7	142.0	143.1	143.8	143.9	144.4	144.9	145.4	146.8	147.3
148.2	151.2	153.3	153.6	153.9	154.5	154.5	155.4	158.4	159.0
163.4	165.9	166.4	166.7	169.9	175.9	177.7			

Таблица 3. Ранжированный ряд для величины E .

Преобразование полученной выборки в вариационный ряд представлено в таблице 3-4.

v	count	v	count	v	count	v	count	v	count
340	1	406	1	441	1	473	1	510	1
344	1	409	1	442	1	475	1	512	1
345	1	414	1	443	1	478	1	514	1
351	2	418	1	448	1	480	2	518	1
352	1	421	2	449	1	481	1	521	1
353	1	422	2	453	1	482	2	522	2
359	1	423	2	454	1	484	1	525	1
362	1	424	1	457	1	488	1	528	1
369	1	426	1	458	1	490	1	541	1
371	2	429	2	459	2	492	1	543	1
386	2	430	1	463	4	493	3	544	1
390	1	431	1	465	1	496	1	562	1
392	1	434	2	467	2	497	1	576	1
393	2	436	2	468	2	501	1	585	1
397	2	437	3	469	1	505	1		
398	1	438	2	471	1	506	1		
400	1	440	1	472	1	508	1		

Таблица 3. Для величины v .

E	count	E	count	E	count	E	count	E	count
71.9	1	109.0	1	121.2	1	134.1	1	145.4	1
82.7	1	110.1	1	121.7	1	134.2	1	146.8	1
84.3	1	110.5	1	121.8	1	134.9	1	147.3	1
85.1	1	111.7	1	121.9	2	135.1	1	148.2	1
86.8	1	112.9	1	122.3	1	135.2	1	151.2	1
87.7	1	113.5	1	122.8	2	136.7	2	153.3	1
89.0	1	114.3	1	122.9	1	137.5	2	153.6	1
89.2	1	114.6	1	124.5	1	137.9	1	153.9	1
91.4	1	115.1	1	125.0	2	139.6	1	154.5	2
91.9	1	115.4	1	126.4	1	139.9	2	155.4	1
95.5	1	115.7	1	126.6	1	140.4	1	158.4	1

E	count	E	count	E	count	E	count	E	count
95.9	1	116.7	1	126.7	1	140.5	1	159.0	1
98.0	1	117.8	1	128.5	1	140.6	1	163.4	1
102.9	1	118.4	1	129.1	1	140.7	1	165.9	1
103.2	1	118.6	1	129.7	1	142.0	1	166.4	1
104.3	1	119.0	1	130.4	1	143.1	1	166.7	1
105.8	1	119.5	1	130.8	1	143.8	1	169.9	1
106.8	1	120.9	1	131.1	2	143.9	1	175.9	1
107.9	1	121.0	1	131.5	1	144.4	1	177.7	1
108.6	1	121.1	1	132.0	1	144.9	1		

Таблица 4. Для величины E.

Преобразование полученной выборки в интервальный вариационный ряд представлено в таблице 5.

Количество интервалов определим по следующей формуле:

$$N = 1 + \text{floor}(\log_2(107)) = 7$$

Ширина интервала: 35.

Интервалы	Частоты	Середины
[340; 375)	12	358
[375; 410)	12	392
[410; 445)	27	428
[445; 480)	23	462
[480; 515)	21	498
[515; 550)	9	532
[550; 585)	3	568

Таблица 5. Для величины v.

Интервалы	Частоты	Середины
[71.9,87)	5	79.5
[87,102)	8	94.6
[102,117)	19	110
[117,132)	32	125

Интервалы	Частоты	Середины
[132,147)	26	140
[147,163)	10	155
[163,178]	7	170

Таблица 5. Для величины E.

Полигон, построенный применительно к интервальному ряду для абсолютных частот представлен на рисунке 1.

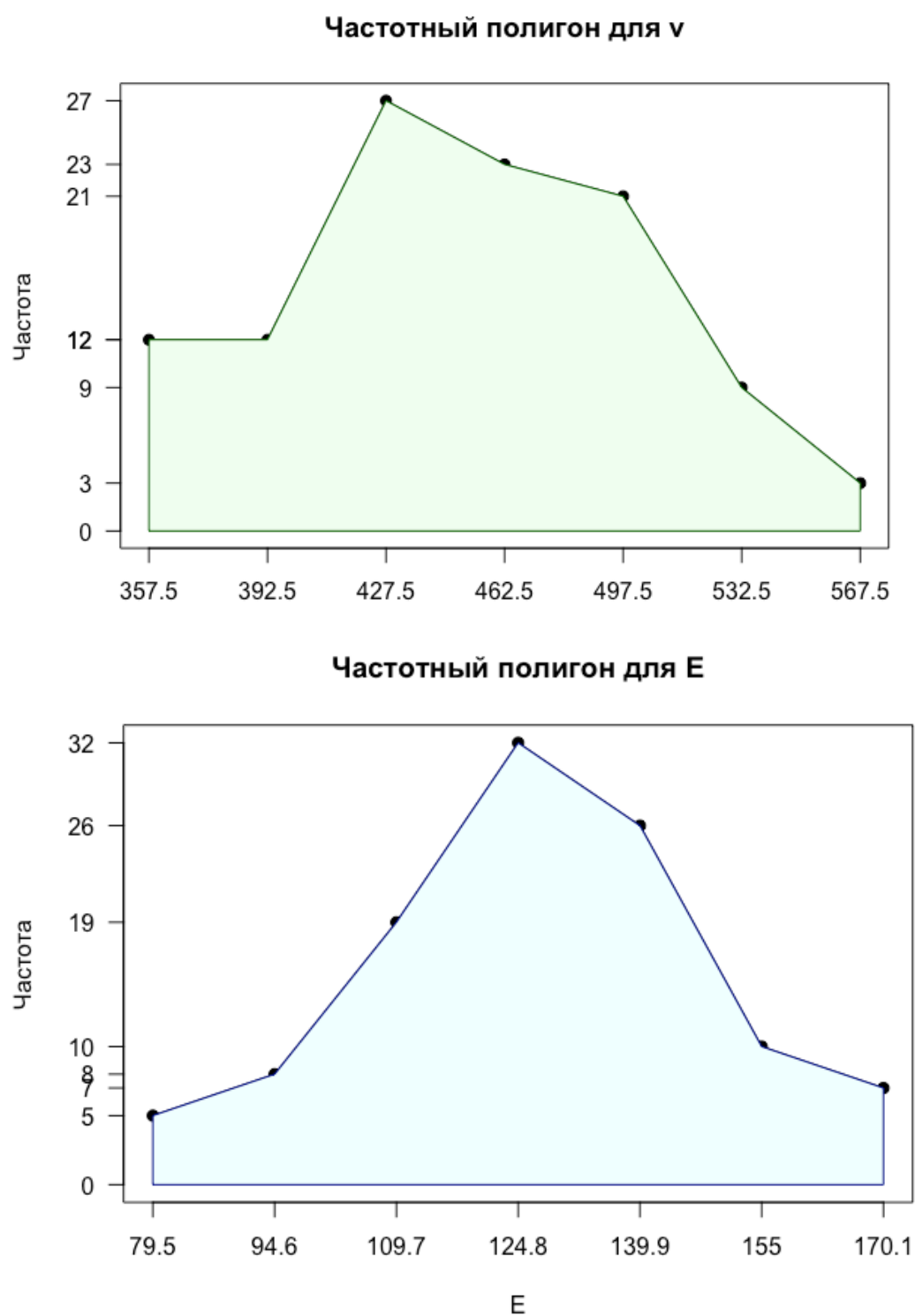


Рис.1

Полигон, построенный применительно к интервальному ряду для относительных частот представлен на рисунке 2.

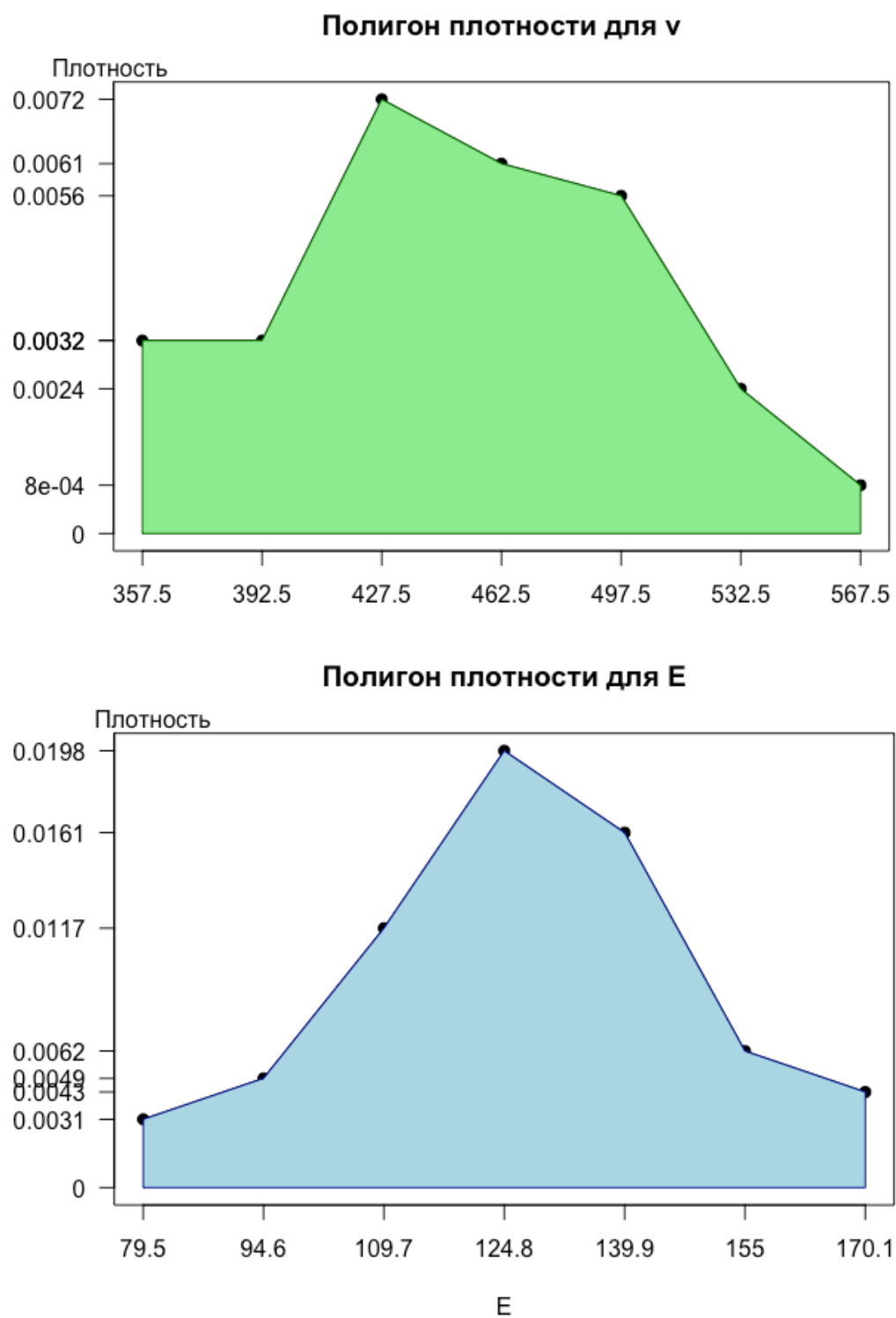


Рис.2

Гистограмма, построенная применительно к интервальному ряду для абсолютных частот представлен на рисунке 3.

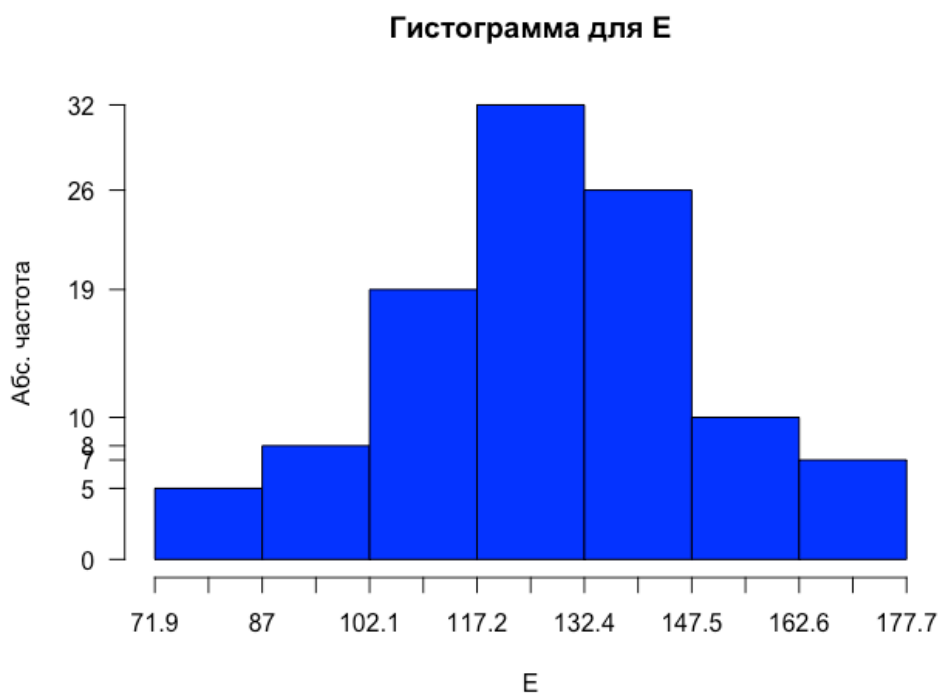
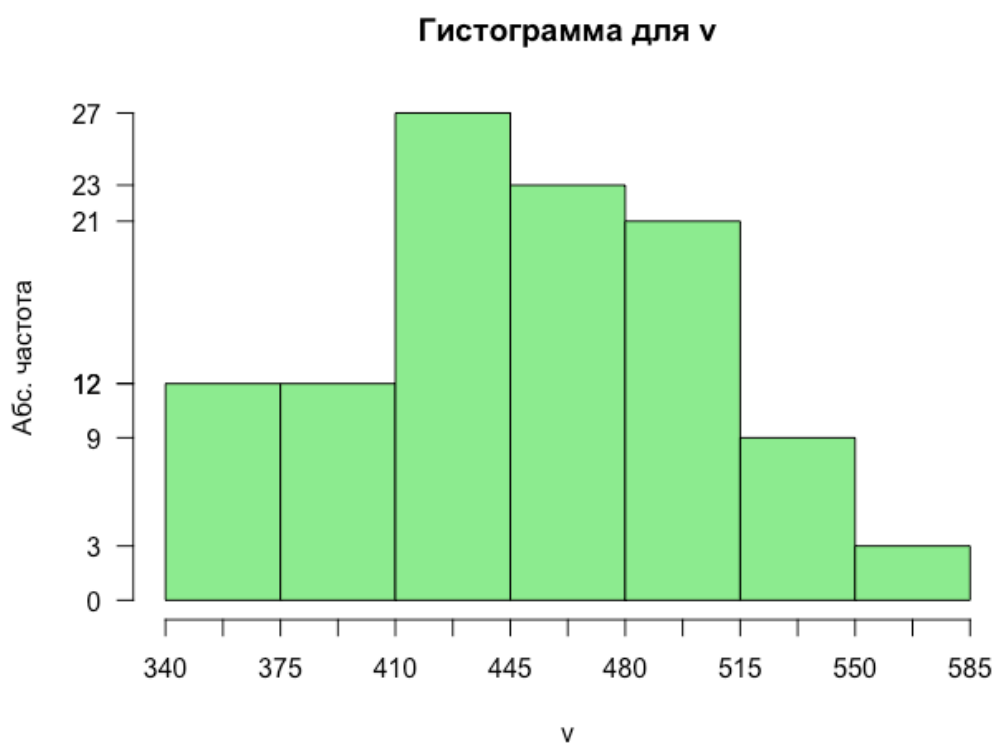


Рис.3

Гистограмма, построенная применительно к интервальному ряду для относительных частот представлен на рисунке 4.

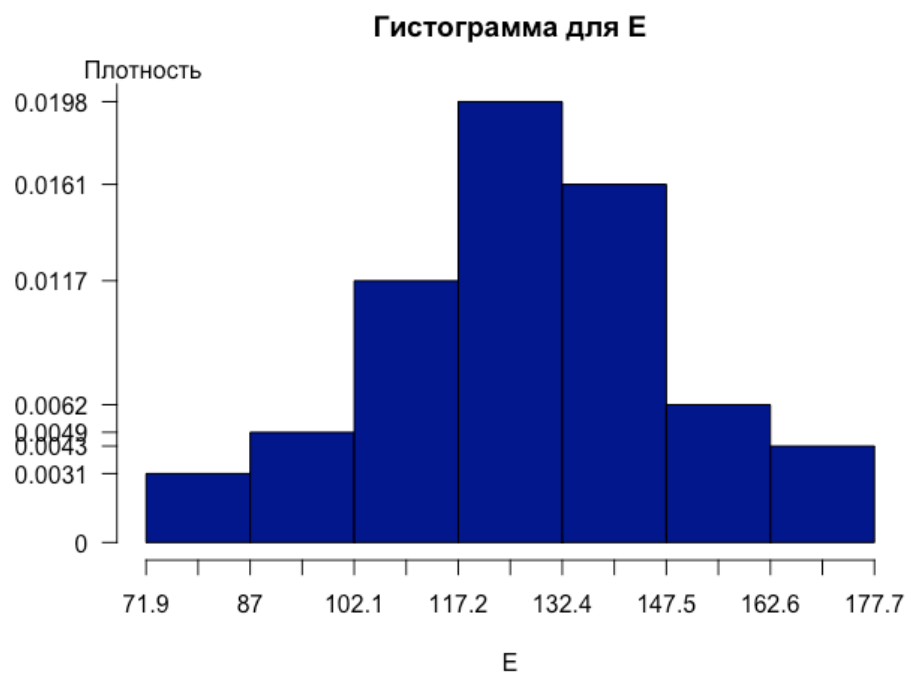
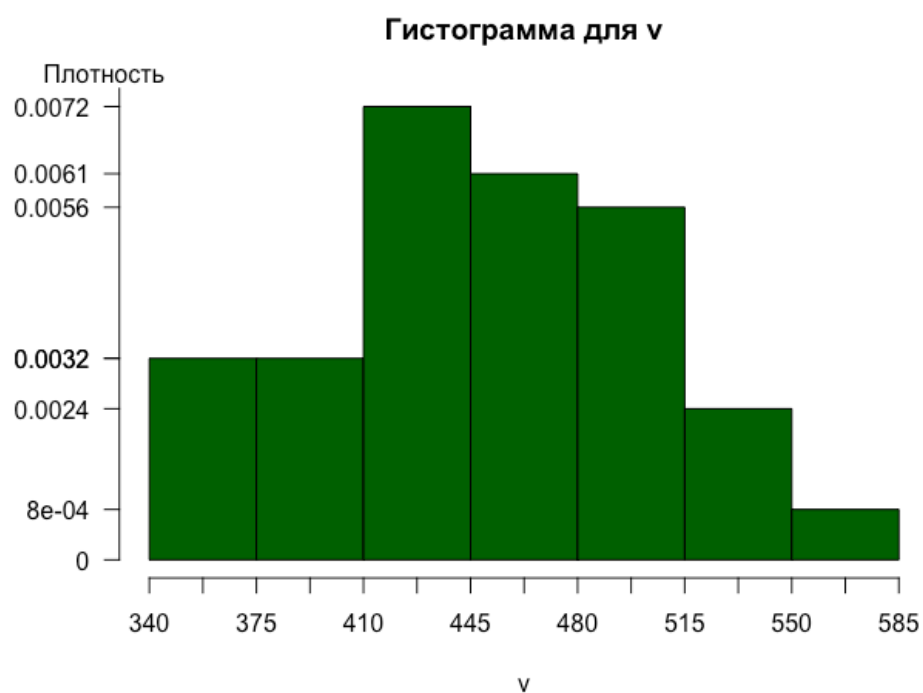


Рис.4

Эмпирическая функция распределения представлена на рисунке 5-6.

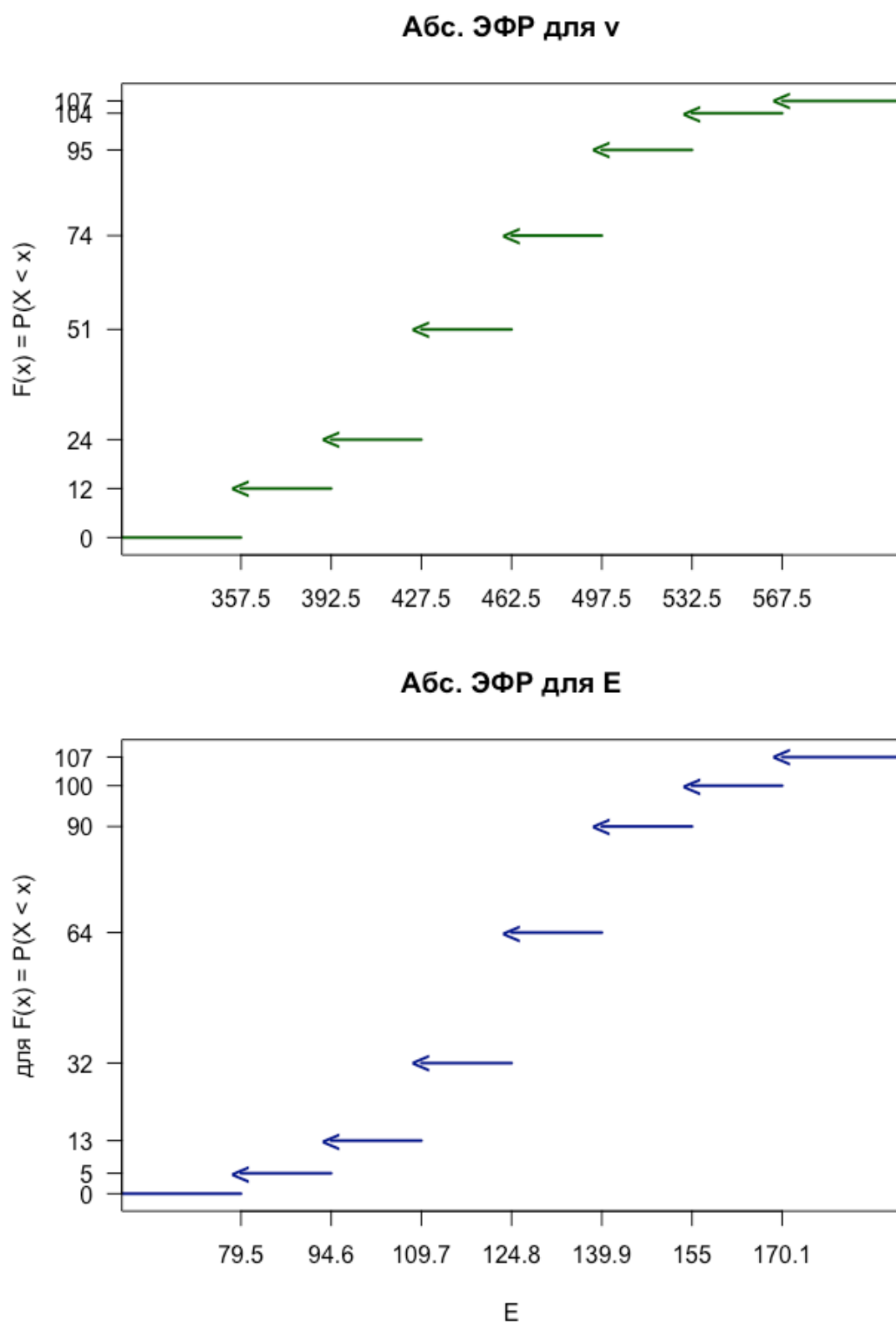
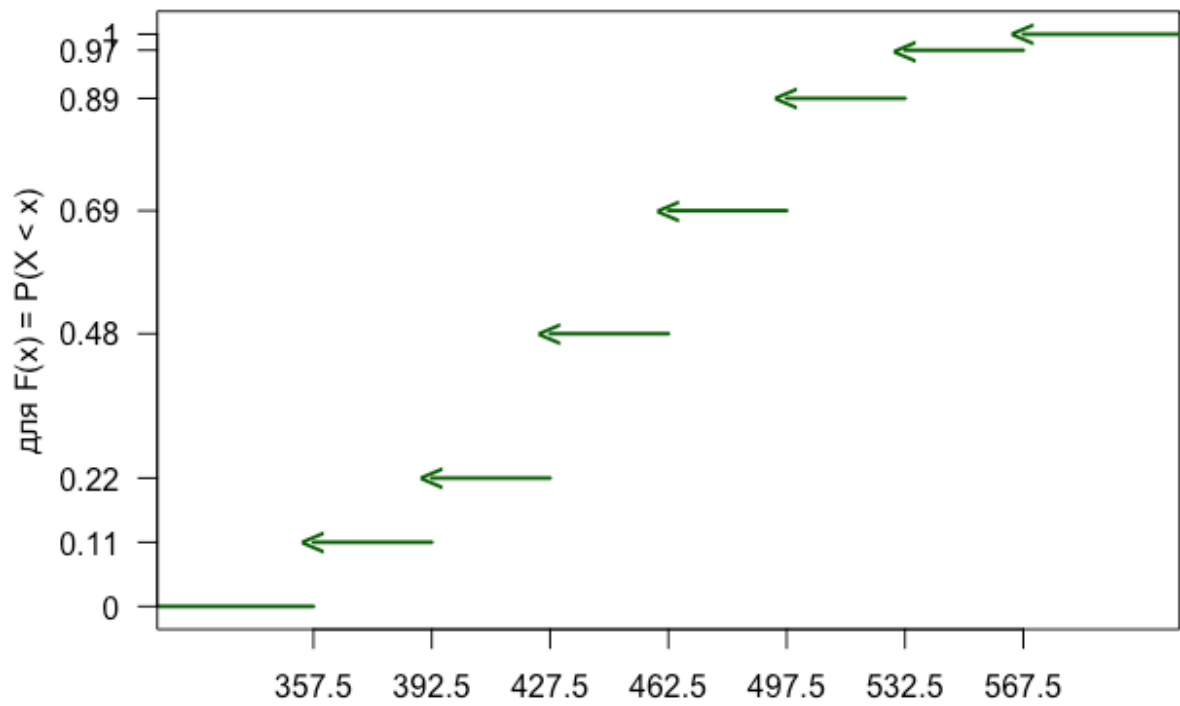


Рис.5

Отн. ЭФР для v



Отн. ЭФР для E

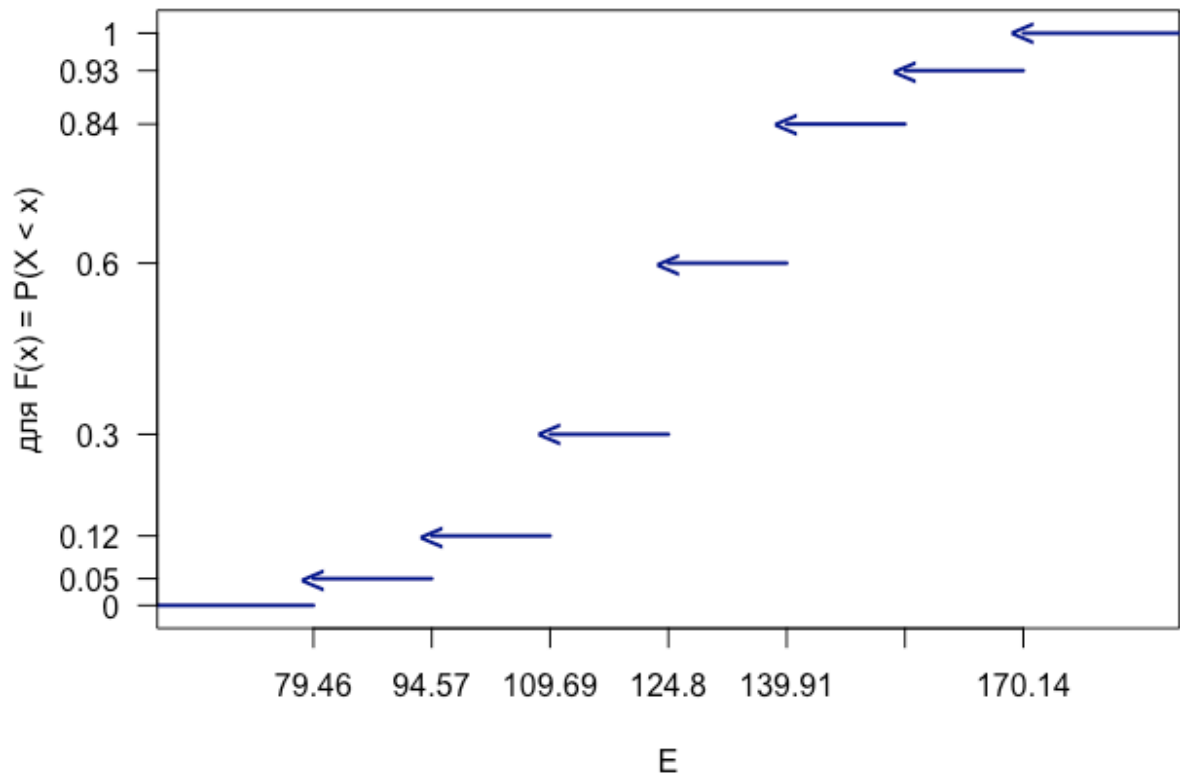


Рис.6

Выводы.

В ходе данной лабораторной работы была сформирована выборка данных и осуществлена её подготовка к статическому анализу.

Выбраны данные для проведения экспериментов. Затем, выборка приведена к ранжированному, вариационному и интервальному видам. Используя полученный интервальный ряд, построен полигон, гистограмма и эмпирическая функция распределения для абсолютных и относительных частот. Из графиков видно, что полученные полигон и гистограмма частот близки к нормальному распределению.

Литература.

1. Серeda В.И. Курс лекций по статическим методам обработки экспериментальных данных. Лекция 1, 23.02.2018
2. Воскобойников Ю.Е. А. Математическая статистика. URL: http://window.edu.ru/resource/305/63305/files/stat_excel.pdf (дата обращения: 30.03.2018)