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

ОТЧЕТ

по лабораторной работе №4
по дисциплине «Статистические методы обработки экспериментальных данных»
Тема: Элементы корреляционного анализа. Проверка статистической гипотезы о равенстве коэффициента корреляции нулю.

Студентка гр. 7381		Алясова А.Н.
Студент гр. 7381		Кортев Ю.В.
Преподаватель		Середа А.-В.И.

Санкт-Петербург
2021

Цель работы.

Освоение основных понятий, связанных с корреляционной зависимостью между случайными величинами, доверительными интервалами, статистическими гипотезами и проверкой их «справедливости».

Основные теоретические положения.

Рассмотрим систему двух случайных величин $\{X; Y\}$. Эти случайные величины могут быть независимыми:

$$f(x, y) = f_1(x) \cdot f_2(y)$$

В противном случае между ними может быть:

a) Функциональная зависимость:

$$y = g(x)$$

b) Статистическая зависимость:

$$\varphi\left(\frac{x}{y}\right) = \frac{f(x, y)}{f_2(y)}$$

$$\phi\left(\frac{y}{x}\right) = \frac{f(x, y)}{f_1(x)}$$

Одним из видов (частным случаем) статистической зависимости является корреляционная зависимость.

Корреляционной называют статистическую зависимость двух случайных величин, при которой изменение значения одной из случайных величин приводит к изменению математического ожидания другой случайной величины (регрессии):

$$M\left(\frac{X}{y}\right) = q_1(y)$$

$$M\left(\frac{Y}{x}\right) = q_2(x)$$

Корреляционный момент:

$$\mu_{xy} = M\{[x - M(X)] \cdot [y - M(Y)]\}$$

Коэффициент корреляции:

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}$$

Для коэффициента корреляции справедливо соотношение:

$$|r_{xy}| \le 1$$

Случайные величины называют коррелированными, если их корреляционный момент или их коэффициент корреляции отличен от нуля. В противном случае эти величины некоррелированные.

Если случайные величины X и Y коррелированы, то они зависимы. Обратное предположение в общем случае неверно:

1. X и Y коррелированы $\Rightarrow$ X и Y зависимы.
2. X и Y некоррелированные $\Leftarrow$ X и Y независимы

Коэффициент корреляции служит мерой тесноты линейной зависимости между случайными величинами X и Y. При $|r_{xy}| = 1$ эта зависимость становится функциональной.

Значение $\bar{r}_{xy}$ – статистической оценки r_{xy} – коэффициента корреляции можно вычислить по формуле:

$$\bar{r}_{xy} = \frac{\sum_{i=1}^{K_y} \sum_{j=1}^{K_x} n_{ij} y_i x_j - N \bar{x}_{\text{в}} \bar{y}_{\text{в}}}{N S_x S_y}$$

При $N > 50$ в случае нормального распределния системы случайных величин $\{X; Y\}$ для оценки значения $\bar{r}_{xy}$ можно использовать соотношение (не является доверительным интервалом):

$$\bar{r}_{xy} - 3\frac{1 - \bar{r}_{xy}^2}{\sqrt{N}} \le r_{xy} \le \bar{r}_{xy} + 3\frac{1 + \bar{r}_{xy}^2}{\sqrt{N}}$$

Распределение $\bar{r}_{xy}$ при определённых условиях можно удовлетворительно аппроксимировать нормальным законом. Однако при увеличении интенсивности связи распределение $\bar{r}_{xy}$ становится всё более ассиметричным.

С помощью преобразования Фишера перейдём к случайной величине z:

$$\bar{z} = 0.5 \ln \frac{1 + \bar{r}_{xy}}{1 - \bar{r}_{xy}}$$

Распределение z при неограниченном возрастании объёма выборки асимптотически нормальное со значением СКВО:

$$\bar{\sigma}_z = \frac{1}{\sqrt{N-3}}$$

В результате доверительный интервал для r_{xy} генеральной совокупности с доверительной вероятностью γ определяется по следующей схеме:

1. По формуле (1) вычисляется выборочное значение $\bar{z}$.
2. По формуле (2) вычисляется значение $\bar{\sigma}_z$.
3. Интервал для генерального значения представляется в виде:

$$(\bar{z} - \lambda(\gamma)\bar{\sigma}_z; \bar{z} + \lambda(\gamma)\bar{\sigma}_z)$$

где значение $\lambda(\gamma)$ должно удовлетворять условию:

$$\Phi\big(\lambda(\gamma)\big) = \frac{\gamma}{2}$$

4. Для пересчёта интервала в доверительных интервал для коэффициента корреляции с тем же значением γ необходимо воспользоваться обратным преобразованием Фишера:

$$r = th(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{e^{2z} - 1}{e^{2z} + 1}$$

Пусть имеется выборка объёма N значений двумерной нормально распределённой случайной величины $\{X; Y\}$ и вычислено значение выборочного коэффициента корреляции $\bar{r}_{xy} \neq 0$. Поскольку $\bar{r}_{xy}$ является случайной величиной, то это ещё не значит, что r_{xy} – коэффициент корреляции для генеральной совокупности тоже отличен от нуля.

Возникает необходимость проверить гипотезу $H_0: r_{xy} = 0$. Альтернативой будет гипотеза $H_1: r_{xy} \neq 0$. Если основная гипотеза отвергается, то это означает, что выборочный коэффициент корреляции $\bar{r}_{xy}$ значимо отличается от нуля (значим). В противном случае $\bar{r}_{xy}$ – незначим.

В качестве критерия проверки статистической гипотезы о значимости выборочного коэффициента корреляции можно принять случайную величину:

$$T = \frac{\bar{r}_{xy}\sqrt{N-2}}{\sqrt{1-\bar{r}_{xy}^2}}$$

При справедливости нулевой гипотезы H_0 случайная величина T распределена по закону Стьюдента с $k = N - 2$ степенями свободы. Критическая область для данного критерия двусторонняя.

Проверка гипотезы осуществляется по стандартной схеме:

1. По формуле (3) вычисляется значение $T_{\text{набл}}$.
2. По заданному уровню значимости α и значению k из таблицы определяется значение $t_{\text{крит}}(\alpha, k)$.
3. Если $|T_{\text{набл}}| \leq t_{\text{крит}}(\alpha, k)$ – нет оснований отвергать гипотезу H_0.

Если $|T_{\text{набл}}| > t_{\text{крит}}(\alpha, k)$ – основная гипотеза H_0 с выборочными данными должна быть отвергнута.

Постановка задачи.

Из заданной генеральной совокупности сформировать выборку по второму признаку. Провести статистическую обработку второй выборки в объеме лабораторных работ №1 и №2, с целью определения точечных статистических оценок параметров распределения исследуемого признака (математического ожидания, дисперсии, среднеквадратичного отклонения, асимметрии и эксцесса). Для системы двух случайных величин X (первый признак) и Y (второй признак) сформировать двумерную выборку и найти статистическую оценку коэффициента корреляции, построить доверительный интервал для коэффициента корреляции и осуществить проверку статистической гипотезы о равенстве коэффициента корреляции нулю. Полученные результаты содержательно проинтерпретировать.

Порядок выполнения работы.

1. Провести статистическую обработку второй выборки в объеме лабораторных работ №1 и №2, с целью определения точечных статистических оценок

параметров распределения исследуемого признака (математического ожидания, дисперсии, среднеквадратичного отклонения, асимметрии, эксцесса, моды и медианы). Оформить результаты в виде таблицы, сделать выводы.

2. Построить двумерный интервальный вариационный ряд, оформить в виде таблицы.

3. По полученному двумерному интервальному вариационному ряду построить корреляционную таблицу, сделать выводы.

4. Исходя из результатов корреляционной таблицы вычислить статистическую оценку корреляционного момента.

5. Вычислить коэффициент корреляции, сделать выводы.

6. Построить доверительный интервал для коэффициента корреляции при уровне значимости $\gamma \in \{0.95, 0.99\}$ сделать выводы.

7. Осуществить проверку статистической гипотезы о равенстве коэффициента корреляции нулю при заданном уровне значимости $\alpha = 0.05$ сделать выводы.

Выполнение работы.

1. Провести статистическую обработку второй выборки в объеме лабораторных работ №1 и №2, с целью определения точечных статистических оценок параметров распределения исследуемого признака (математического ожидания, дисперсии, среднеквадратичного отклонения, асимметрии, эксцесса, моды и медианы). Оформить результаты в виде таблицы, сделать выводы.

Результаты формирования второй выборки заданного объема из имеющейся генеральной совокупности экспериментальных данных представлены в табл. 1-2.

Объём выборки: 117.

Таблица 1

№	***nu***	***E***	№	***nu***	***E***	№	***nu***	***E***	№	***nu***	***E***	№	***nu***	***E***
1	480	153,3	25	408	110,0	49	405	103,6	73	465	127,7	97	487	146,0
2	510	129,4	26	331	74,1	50	434	140,4	74	390	108,1	98	532	158,7
3	426	119,0	27	467	113,0	51	344	86,8	75	463	129,2	99	330	71,1
4	482	139,9	28	545	145,3	52	415	119,7	76	468	128,9	100	438	134,1
5	393	103,2	29	396	83,8	53	463	136,7	77	488	134,1	101	593	187,4
6	510	162,3	30	351	102,9	54	475	143,6	78	443	137,4	102	445	124,7
7	403	123,9	31	503	148,5	55	463	144,9	79	505	155,8	103	518	154,0
8	506	158,4	32	402	120,8	56	392	82,7	80	395	109,1	104	496	141,7
9	393	122,8	33	542	146,1	57	452	140,5	81	474	132,5	105	473	136,4
10	442	115,4	34	437	124,3	58	504	143,8	82	490	139,9	106	522	154,5
11	411	112,9	35	453	119,5	59	443	122,9	83	396	90,1	107	547	154,7
12	514	153,6	36	386	105,8	60	461	138,6	84	362	97,9	108	560	169,8
13	525	156,5	37	434	122,3	61	340	85,1	85	566	175,7	109	412	127,8
14	543	155,4	38	418	118,4	62	438	134,9	86	418	109,3	110	444	130,0
15	412	116,3	39	391	107,5	63	523	148,7	87	502	132,5	111	437	121,8
16	449	124,5	40	399	100,0	64	416	120,5	88	500	155,5	112	462	138,8
17	482	136,4	41	486	139,4	65	483	143,4	89	359	71,9	113	438	122,2
18	569	157,4	42	421	124,2	66	440	128,5	90	443	135,7	114	406	110,1
19	484	147,5	43	496	143,1	67	423	131,1	91	421	118,0	115	413	106,7

20	472	134,2	44	463	121,2	68	386	95,5	92	433	128,2	116	458	121,7
21	453	124,2	45	508	159,0	69	321	86,1	93	514	174,6	117	408	117,0
22	422	117,9	46	419	105,3	70	433	131,5	94	320	72,6			
23	320	64,5	47	434	108,7	71	351	89,0	95	406	113,8			
24	547	164,4	48	440	126,7	72	481	148,3	96	465	140,9			

Таблица 2

i	x_i	i	x_i	i	x_i	i	x_i	i	x_i	i	x_i	i	x_i
1	153,3	18	157,4	35	119,5	52	119,7	69	86,1	86	109,3	103	154,0
2	129,4	19	147,5	36	105,8	53	136,7	70	131,5	87	132,5	104	141,7
3	119,0	20	134,2	37	122,3	54	143,6	71	89,0	88	155,5	105	136,4
4	139,9	21	124,2	38	118,4	55	144,9	72	148,3	89	71,9	106	154,5
5	103,2	22	117,9	39	107,5	56	82,7	73	127,7	90	135,7	107	154,7
6	162,3	23	64,5	40	100,0	57	140,5	74	108,1	91	118,0	108	169,8
7	123,9	24	164,4	41	139,4	58	143,8	75	129,2	92	128,2	109	127,8
8	158,4	25	110,0	42	124,2	59	122,9	76	128,9	93	174,6	110	130,0
9	122,8	26	74,1	43	143,1	60	138,6	77	134,1	94	72,6	111	121,8
10	115,4	27	113,0	44	121,2	61	85,1	78	137,4	95	113,8	112	138,8
11	112,9	28	145,3	45	159,0	62	134,9	79	155,8	96	140,9	113	122,2
12	153,6	29	83,8	46	105,3	63	148,7	80	109,1	97	146,0	114	110,1
13	156,5	30	102,9	47	108,7	64	120,5	81	132,5	98	158,7	115	106,7
14	155,4	31	148,5	48	126,7	65	143,4	82	139,9	99	71,1	116	121,7
15	116,3	32	120,8	49	119,5	66	128,5	83	90,1	100	134,1	117	117,0
16	124,5	33	146,1	50	105,8	67	131,1	84	97,9	101	187,4		
17	136,4	34	124,3	51	122,3	68	95,5	85	175,7	102	124,7		

Преобразование полученной выборки в ранжированный ряд представлено в табл. 3.

Таблица 3

i	x_i	i	x_i	i	x_i	i	x_i	i	x_i	i	x_i	i	x_i
1	64,5	18	103,6	35	117,9	52	124,2	69	134,1	86	141,7	103	154,7
2	71,1	19	105,3	36	118,0	53	124,3	70	134,1	87	143,1	104	155,4
3	71,9	20	105,8	37	118,4	54	124,5	71	134,2	88	143,4	105	155,5
4	72,6	21	106,7	38	119,0	55	124,7	72	134,9	89	143,6	106	155,8
5	74,1	22	107,5	39	119,5	56	126,7	73	135,7	90	143,8	107	156,5

6	82,7	23	108,1	40	119,7	57	127,7	74	136,4	91	144,9	108	157,4
7	83,8	24	108,7	41	120,5	58	127,8	75	136,4	92	145,3	109	158,4
8	85,1	25	109,1	42	120,8	59	128,2	76	136,7	93	146,0	110	158,7
9	86,1	26	109,3	43	121,2	60	128,5	77	137,4	94	146,1	111	159,0
10	86,8	27	110,0	44	121,7	61	128,9	78	138,6	95	147,5	112	162,3
11	89,0	28	110,1	45	121,8	62	129,2	79	138,8	96	148,3	113	164,4
12	90,1	29	112,9	46	122,2	63	129,4	80	139,4	97	148,5	114	169,8
13	95,5	30	113,0	47	122,3	64	130,0	81	139,9	98	148,7	115	174,6
14	97,9	31	113,8	48	122,8	65	131,1	82	139,9	99	153,3	116	175,7
15	100,0	32	115,4	49	122,9	66	131,5	83	140,4	100	153,6	117	187,4
16	102,9	33	116,3	50	123,9	67	132,5	84	140,5	101	154,0		
17	103,2	34	117,0	51	124,2	68	132,5	85	140,9	102	154,5		

Из табл. 3 можно увидеть, что наименьшее значение в выборке $x_{min} = 64{,}5$, а наибольшее значение $x_{max} = 187{,}4$.

Преобразование полученной выборки в вариационный ряд с абсолютными частотами представлено в табл. 4.

Таблица 4

i	x_i	p_i	i	x_i	p_i	i	x_i	p_i	i	x_i	p_i	i	x_i	p_i
1	64,5	1	24	108,7	1	47	122,3	1	70	135,7	1	93	148,7	1
2	71,1	1	25	109,1	1	48	122,8	1	71	136,4	2	94	153,3	1
3	71,9	1	26	109,3	1	49	122,9	1	72	136,7	1	95	153,6	1
4	72,6	1	27	110,0	1	50	123,9	1	73	137,4	1	96	154,0	1
5	74,1	1	28	110,1	1	51	124,2	2	74	138,6	1	97	154,5	1
6	82,7	1	29	112,9	1	52	124,3	1	75	138,8	1	98	154,7	1
7	83,8	1	30	113,0	1	53	124,5	1	76	139,4	1	99	155,4	1
8	85,1	1	31	113,8	1	54	124,7	1	77	139,9	2	100	155,5	1
9	86,1	1	32	115,4	1	55	126,7	1	78	140,4	1	101	155,8	1
10	86,8	1	33	116,3	1	56	127,7	1	79	140,5	1	102	156,5	1
11	89,0	1	34	117,0	1	57	127,8	1	80	140,9	1	103	157,4	1
12	90,1	1	35	117,9	1	58	128,2	1	81	141,7	1	104	158,4	1
13	95,5	1	36	118,0	1	59	128,5	1	82	143,1	1	105	158,7	1
14	97,9	1	37	118,4	1	60	128,9	1	83	143,4	1	106	159,0	1
15	100,0	1	38	119,0	1	61	129,2	1	84	143,6	1	107	162,3	1

16	102,9	1	39	119,5	1	62	129,4	1	85	143,8	1	108	164,4	1
17	103,2	1	40	119,7	1	63	130,0	1	86	144,9	1	109	169,8	1
18	103,6	1	41	120,5	1	64	131,1	1	87	145,3	1	110	174,6	1
19	105,3	1	42	120,8	1	65	131,5	1	88	146,0	1	111	175,7	1
20	105,8	1	43	121,2	1	66	132,5	2	89	146,1	1	112	187,4	1
21	106,7	1	44	121,7	1	67	134,1	2	90	147,5	1			
22	107,5	1	45	121,8	1	68	134,2	1	91	148,3	1			
23	108,1	1	46	122,2	1	69	134,9	1	92	148,5	1			

Преобразование полученной выборки в вариационный ряд с относительными частотами представлено в табл. 5.

Таблица 5

i	x_i	p_i	i	x_i	p_i	i	x_i	p_i	i	x_i	p_i
1	64,5	0,008547	29	112,9	0,008547	57	127,8	0,008547	85	143,8	0,008547
2	71,1	0,008547	30	113,0	0,008547	58	128,2	0,008547	86	144,9	0,008547
3	71,9	0,008547	31	113,8	0,008547	59	128,5	0,008547	87	145,3	0,008547
4	72,6	0,008547	32	115,4	0,008547	60	128,9	0,008547	88	146,0	0,008547
5	74,1	0,008547	33	116,3	0,008547	61	129,2	0,008547	89	146,1	0,008547
6	82,7	0,008547	34	117,0	0,008547	62	129,4	0,008547	90	147,5	0,008547
7	83,8	0,008547	35	117,9	0,008547	63	130,0	0,008547	91	148,3	0,008547
8	85,1	0,008547	36	118,0	0,008547	64	131,1	0,008547	92	148,5	0,008547
9	86,1	0,008547	37	118,4	0,008547	65	131,5	0,008547	93	148,7	0,008547
10	86,8	0,008547	38	119,0	0,008547	66	132,5	0,017094	94	153,3	0,008547
11	89,0	0,008547	39	119,5	0,008547	67	134,1	0,017094	95	153,6	0,008547
12	90,1	0,008547	40	119,7	0,008547	68	134,2	0,008547	96	154,0	0,008547
13	95,5	0,008547	41	120,5	0,008547	69	134,9	0,008547	97	154,5	0,008547
14	97,9	0,008547	42	120,8	0,008547	70	135,7	0,008547	98	154,7	0,008547
15	100,0	0,008547	43	121,2	0,008547	71	136,4	0,017094	99	155,4	0,008547
16	102,9	0,008547	44	121,7	0,008547	72	136,7	0,008547	100	155,5	0,008547
17	103,2	0,008547	45	121,8	0,008547	73	137,4	0,008547	101	155,8	0,008547
18	103,6	0,008547	46	122,2	0,008547	74	138,6	0,008547	102	156,5	0,008547
19	105,3	0,008547	47	122,3	0,008547	75	138,8	0,008547	103	157,4	0,008547
20	105,8	0,008547	48	122,8	0,008547	76	139,4	0,008547	104	158,4	0,008547
21	106,7	0,008547	49	122,9	0,008547	77	139,9	0,017094	105	158,7	0,008547

22	107,5	0,008547	50	123,9	0,008547	78	140,4	0,008547	106	159,0	0,008547
23	108,1	0,008547	51	124,2	0,017094	79	140,5	0,008547	107	162,3	0,008547
24	108,7	0,008547	52	124,3	0,008547	80	140,9	0,008547	108	164,4	0,008547
25	109,1	0,008547	53	124,5	0,008547	81	141,7	0,008547	109	169,8	0,008547
26	109,3	0,008547	54	124,7	0,008547	82	143,1	0,008547	110	174,6	0,008547
27	110,0	0,008547	55	126,7	0,008547	83	143,4	0,008547	111	175,7	0,008547
28	110,1	0,008547	56	127,7	0,008547	84	143,6	0,008547	112	187,4	0,008547

Для определения количества интервалов используем формулу Стерджесса:

$$k = 1 + 3{,}322 * \log(n),$$

где n – объем выборки.

Используя в качестве $n = 117$, получаем, что $k = 8$.

Чтобы определить шаг, с которым формировать интервалы, использована формула:

$$h = \frac{x_{max} - x_{min}}{k}.$$

Соответственно, для $x_{min} = 64{,}5$, $x_{max} = 187{,}4$ и $k = 8$ получаем, что $h \approx 15$. Полученный интервальный ряд приведен в табл. 6.

Таблица 6

Интервал	**Абсолютная частота**	**Относительная частота**
[64,5 ; 79,5)	5	0,042735
[79,5 ; 94,5)	7	0,059829
[94,5 ; 109,5)	14	0,119658
[109,5 ; 124,5)	27	0,230769
[124,5 ; 139,5)	27	0,230769
[139,5 ; 154,5)	21	0,179487
[154,5 ; 169,5)	12	0,102564
[169,5 ; 184,5)	3	0,025641
[184,5 ; 187,4]	1	0,008547

В сумме абсолютные частоты дают 117, что соответствует объему выборки, а относительные частоты суммируются к единице.

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

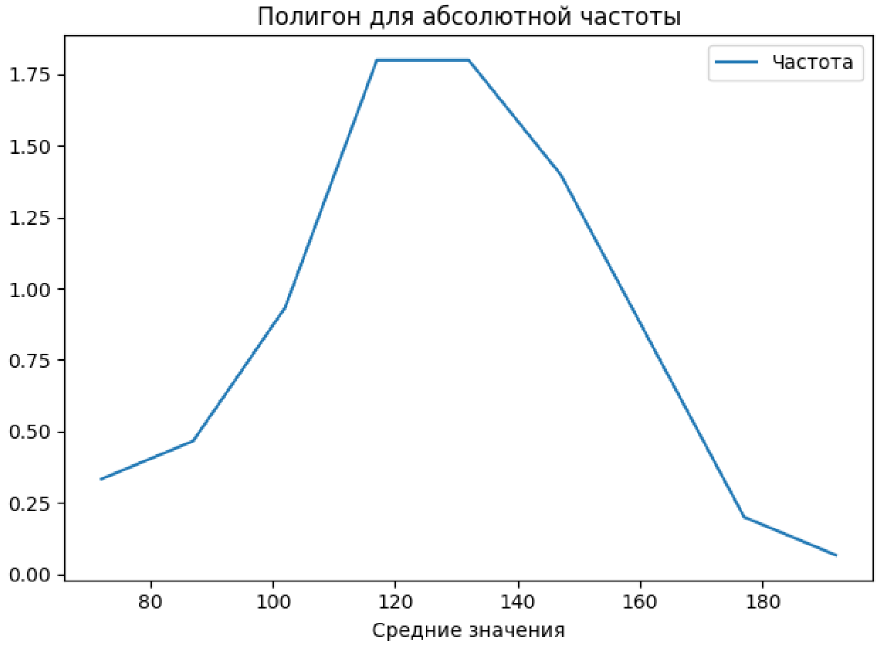

Рисунок 1 – Полигон для абсолютной частоты

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

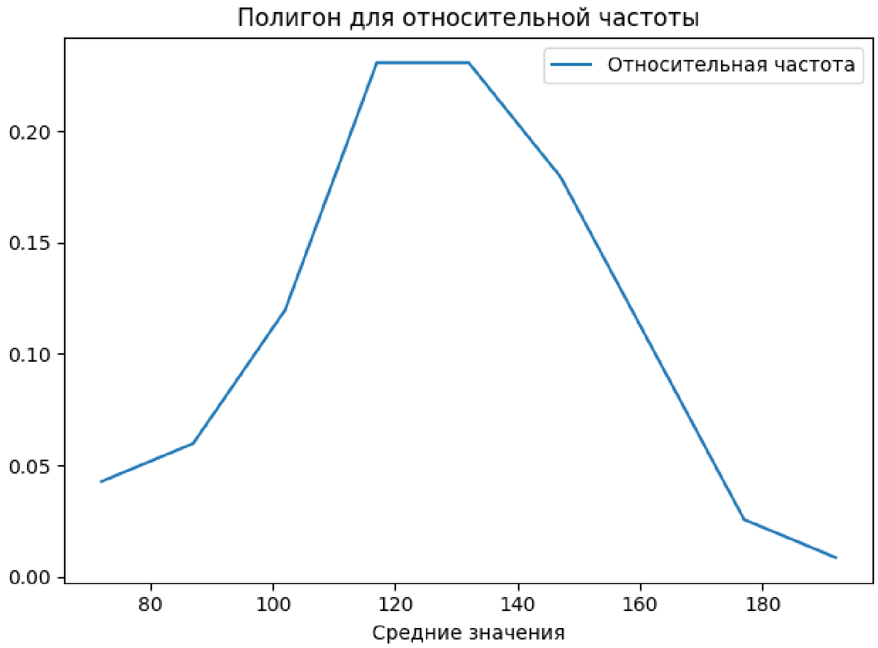

Рисунок 2 – Полигон для относительной частоты

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

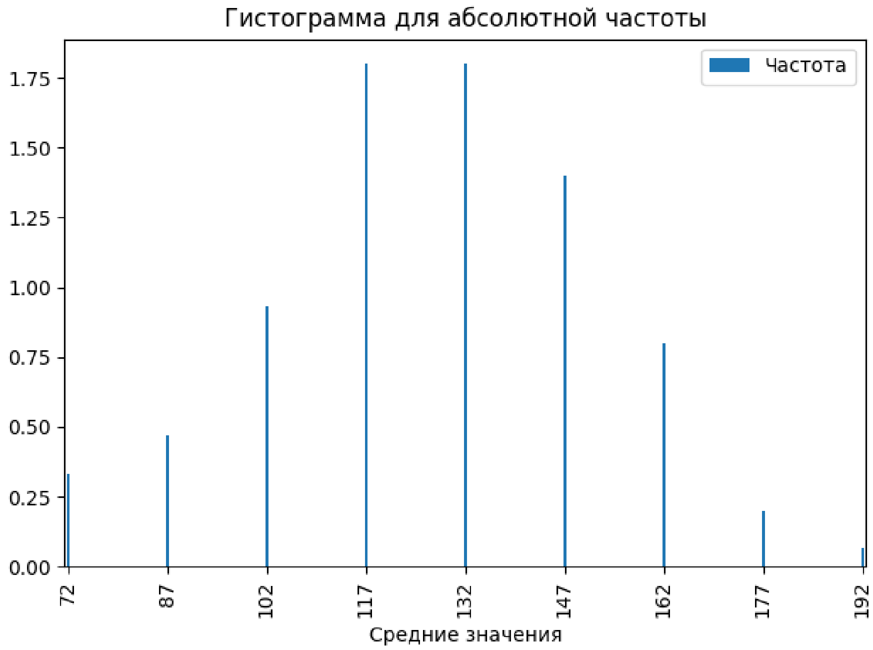

Рисунок 3 – Гистограмма для абсолютной частоты

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

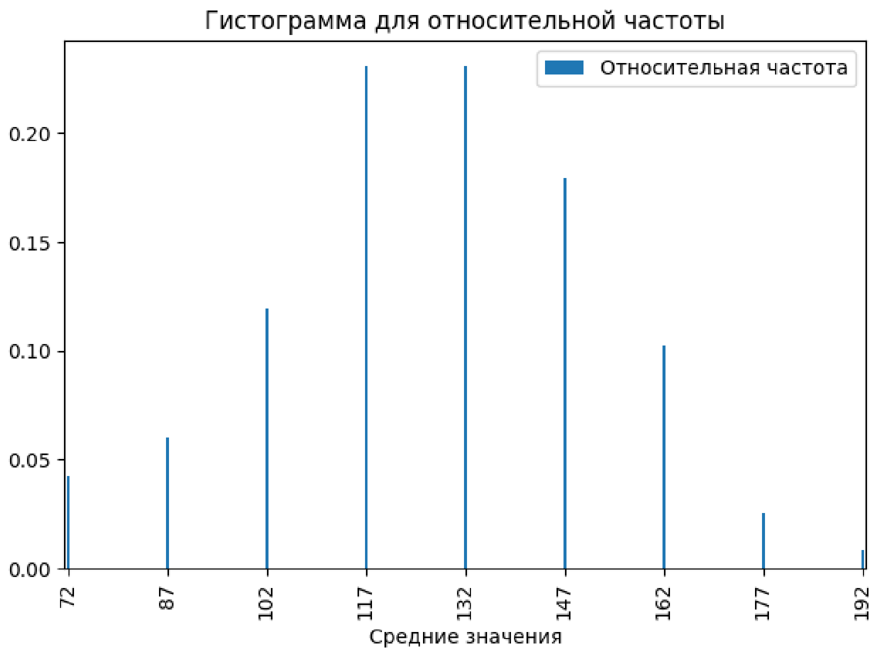

Рисунок 4 – Гистограмма для относительной частоты

Эмпирическая функция распределения, построенная применительно к интервальному ряду для абсолютных частот представлен на рис. 5.

Функция распределения:

$$F(72) = 0{,}042735042735042736$$

$$F(87) = 0{,}10256410256410256$$

$$F(102) = 0{,}22222222222222224$$

$$F(117) = 0{,}45299145299145305$$

$$F(132) = 0{,}6837606837606836$$

$$F(147) = 0{,}8632478632478631$$

$$F(162) = 0{,}9658119658119655$$

$$F(177) = 0{,}991452991452991$$

$$F(192) = 0{,}9999999999999999$$

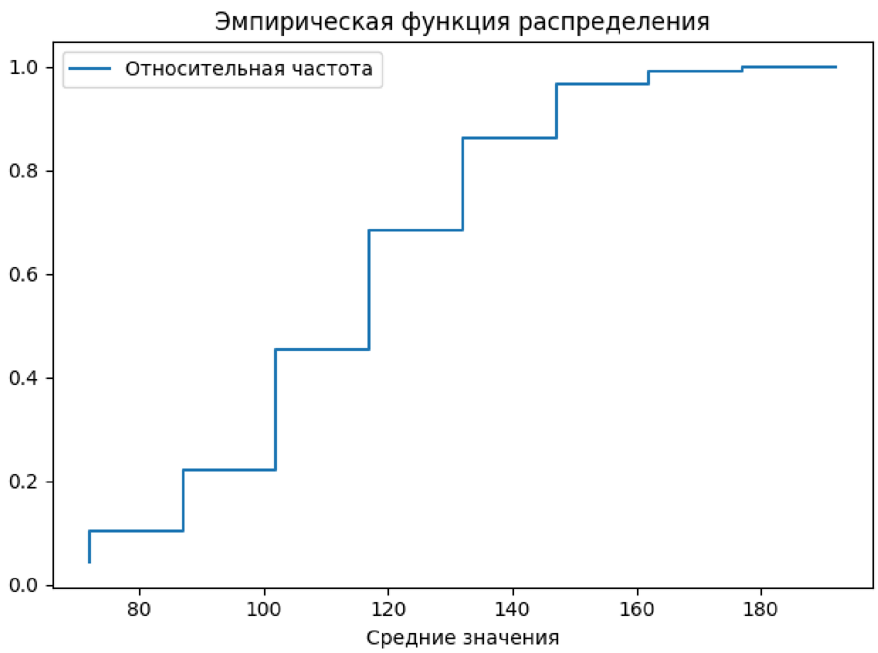

Рисунок 5 – График эмпирической функции распределения

Найдем условные моменты по формуле:

$$\widetilde{M}_l = \frac{1}{N}\sum_{i=1}^{k} \tilde{x}_i^l n_i,$$

$$\tilde{x}_i = \frac{1}{h}(x_i - C),$$

где h – длина интервала, $C = x_5$ – ложный ноль.

Результаты вычислений представлены в табл. 7.

Таблица 7

x_i	n_i	$\tilde{x}_i$	$\tilde{x}_i * n_i$	$\tilde{x}_i^2 * n_i$	$\tilde{x}_i^3 * n_i$	$\tilde{x}_i^4 * n_i$	$\widetilde{(x_i^4} + 1)\text{^}4 * n_i$
72	5	-4	-20	80	-320	1280	405
87	7	-3	-21	63	-189	567	112
102	14	-2	-28	56	-112	224	14
117	27	-1	-27	27	-27	27	0
132	27	0	0	0	0	0	27
147	21	1	21	21	21	21	336
162	12	2	24	48	96	192	972
177	3	3	9	27	81	243	768
192	1	4	4	16	64	256	625
$\sum =$ 1188	$\sum =$ 117	$\sum =$ 0	$\sum =$ -38	$\sum =$ 338	$\sum =$ -386	$\sum =$ 2 810	$\sum =$ 3 259
Условные моменты:			$-0{,}3248$	2,8889	$-3{,}2991$	24,0171	

Проверим правильность вычислений:

$$\sum \tilde{x}_i^4 * n_i + 4 * \sum \tilde{x}_i^3 * n_i + 6 * \sum \tilde{x}_i^2 * n_i + 4 * \sum \tilde{x}_i * n_i + \sum n_i = 3259$$

Вычислим статистические оценки математического ожидания:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{k} x_i n_i = 127{,}1282$$

Вычислим статистические оценки дисперсии:

$$D_B = \frac{1}{N} \sum_{i=1}^{k} (x_i - \bar{x})^2 \, n_i = 626{,}2656$$

Отсюда следует, что среднеквадратическое отклонение:

$$\sigma = \sqrt{D_{\text{В}}} = \sqrt{626{,}2656} = 25{,}0253$$

Найдем исправленную выборочную дисперсию:

$$S^2 = \frac{N}{N-1} D_B = \frac{117}{116} * 626{,}2656 = 631{,}6645$$

$$S = \sqrt{S^2} = \sqrt{631{,}6645} = 25{,}1329$$

Для вычисления ассиметрии и эксцесса найдем центральные эмпирические моменты третьего и четвертого порядка:

$$m_3 = \left(\widetilde{M}_3 - 3\widetilde{M}_2\widetilde{M}_1 + 2\widetilde{M}_1^3\right) * h^3 = -1865{,}8735$$

$$m_4 = \left(\widetilde{M}_4 - 4\widetilde{M}_3\widetilde{M}_1 + 6\widetilde{M}_2 * \widetilde{M}_1^2 - 3\widetilde{M}_1^4\right) * h^4 = 1089757{,}2755$$

Вычислим ассиметрию:

$$As = \frac{m_3}{S^3} = \frac{-1865{,}8735}{25{,}1329^3} = -0{,}1175$$

Вычислим эксцесс:

$$\mathrm{Ex} = \frac{m_4}{S^4} - 3 = -0{,}2688$$

Далее найдем моду вариационного ряда по формуле:

$$M_O(X) = x_{M_O} + h\frac{(m_2 - m_1)}{(m_2 - m_1) + (m_2 - m_3)}$$

$$M_O = 124{,}5$$

Далее найдем медиану вариационного ряда по формуле:

$$M_e(X) = x_{M_e} + h\frac{0{,}5n - SM_{e^{-1}}}{n_{M_e}}$$

$$M_e = 127{,}5556$$

2. Построить двумерный интервальный вариационный ряд, оформить в виде таблицы.

Построим двумерный интервальный вариационный ряд (табл. 8).

Таблица 8 - Двумерный интервальный вариационный ряд

y_i \ x_j	[64,5; 79,5)	[79,5; 94,5)	[94,5; 109,5)	[109,5; 124,5)	[124,5; 139,5)	[139,5; 154,5)	[154,5; 169,5)	[169,5; 184,5)	[184,5; 187,4]	N
[320,354)	4	4	1	0	0	0	0	0	0	9
[354,388)	1	0	3	0	0	0	0	0	0	4
[388,422)	0	3	9	14	1	0	0	0	0	27
[422,456)	0	0	1	10	12	2	0	0	0	25
[456,490)	0	0	0	3	12	9	0	0	0	24
[490,524)	0	0	0	0	2	8	6	1	0	17
[524,558)	0	0	0	0	0	2	5	0	0	7

[558,592)	0	0	0	0	0	0	1	2	0	3
[592,593]	0	0	0	0	0	0	0	0	1	1
N	5	7	14	27	27	21	12	3	1	117

3. По полученному двумерному интервальному вариационному ряду построить корреляционную таблицу, сделать выводы.

Основываясь на двумерном интервальном вариационном ряде построим корреляционную таблицу. Результат представлен в табл. 9.

Таблица 9 - Корреляционная таблица

u \ v	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	n_v
-4	4	4	1	0	0	0	0	0	0	9
-3	1	0	3	0	0	0	0	0	0	4
-2	0	3	9	14	1	0	0	0	0	27
-1	0	0	1	10	12	2	0	0	0	25
0	0	0	0	3	12	9	0	0	0	24
1	0	0	0	0	2	8	6	1	0	17
2	0	0	0	0	0	2	5	0	0	7
3	0	0	0	0	0	0	1	2	0	3
4	0	0	0	0	0	0	0	0	1	1
n_u	5	7	14	27	27	21	12	3	1	117

4. Исходя из результатов корреляционной таблицы вычислить статистическую оценку корреляционного момента.

С помощью корреляционной таблицы вычислим статистическую оценку корреляционного момента по формуле:

$$\overline{r_{xy}} = \frac{\sum_{i=1}^{K_y}\sum_{j=1}^{K_x} n_{ij}u_i v_j - N\overline{v_в}\,\overline{u_в}}{NS_v S_u},$$

где $\overline{v_в} = -0{,}7094,\ \overline{u_в} = -0{,}3248$ - условные средние для условных вариант, $S_v = 1{,}7174, S_u = 1{,}6755$ – оценки стандартных отклонений условных вариант.

Для вычисления построим вспомогательную таблицу. Результаты представлены в табл. 10.

Таблица 10 – Вспомогательная таблица для вычисления статистической оценки коэффициента корреляции

	-4	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**N**
-4	64	48	8	0	0	0	0	0	0	120
-3	12	0	18	0	0	0	0	0	0	30
-2	0	18	36	28	0	0	0	0	0	82
-1	0	0	2	10	0	-2	0	0	0	10
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	8	12	3	0	23
2	0	0	0	0	0	4	20	0	0	24
3	0	0	0	0	0	0	6	18	0	24
4	0	0	0	0	0	0	0	0	16	16
N	76	66	64	38	0	10	38	21	16	329

$$\overline{r_{xy}} = \frac{\sum_{i=1}^{K_y}\sum_{j=1}^{K_x} n_{ij}u_i v_j - N\overline{v_{\text{в}}}\,\overline{u_{\text{в}}}}{NS_vS_u} = 0{,}8972$$

5. Вычислить коэффициент корреляции, сделать выводы.

Оценим значение r_{xy} для случая нормального распределения по формуле:

$$\overline{r_{xy}} - 3\frac{1+0{,}8972^2}{\sqrt{N}} \le r_{xy} \le \overline{r_{xy}} + 3\frac{1+\overline{r_{xy}}^2}{\sqrt{N}}$$

$$0{,}8972 - 3\frac{1+0{,}8972^2}{\sqrt{117}} \le r_{xy} \le 0{,}8972 + 3\frac{1+0{,}8972^2}{\sqrt{117}}$$

$$0{,}8972 - 3\frac{1+0{,}8972^2}{\sqrt{117}} \le r_{xy} \le 0{,}8972 + 3\frac{1+0{,}8972^2}{\sqrt{117}}$$

$$0{,}3966 \le r_{xy} \le 1{,}3978$$

Получили коэффициент корреляции r_{xy} отличный от нуля, а значит случайные величины в выборке коррелированы и зависимы.

6. Построить доверительный интервал для коэффициента корреляции при уровне значимости $\boldsymbol{\gamma\epsilon\{0,95;0,99\}}$ сделать выводы.

Перейдем к случайной величине z:

$$\bar{z} = 0{,}5\ln\frac{1+\overline{r_{xy}}}{1-\overline{r_{xy}}} = 0{,}5\ln\frac{1+0{,}8972}{1-0{,}8972} = 1{,}4577$$

Вычислим СКВО для распределения z:

$$\overline{\sigma_z} = \frac{1}{\sqrt{N-3}} = 0{,}0937$$

Доверительный интервал для $\bar{z}$ с доверительной вероятностью γ будет определяться:

$$z\epsilon(\bar{z} - \lambda(\gamma)\overline{\sigma_z}\ ;\ \bar{z} - \lambda(\gamma)\overline{\sigma_z}\ ),$$

где $\lambda(\gamma)$ должно удовлетворять условию:

$$\Phi[\lambda(\gamma)] = \frac{\gamma}{2}.$$

Тогда для $\gamma = 0{,}95$, $\lambda(\gamma) = 1{,}96$:

$$z\ \epsilon\ (1{,}4577 - 1{,}96 * 0{,}0937\ ;\ 1{,}4577 + 1{,}96 * 0{,}0937)$$

$$z\ \epsilon\ (1{,}2740\ ; 1{,}6413\ )$$

Тогда для $\gamma = 0{,}99$, $\lambda(\gamma) = 2{,}58$:

$$z\ \epsilon\ (1{,}4577 - 2{,}58 * 0{,}0937\ ;\ 1{,}4577 + 2{,}58 * 0{,}0937)$$

$$z\ \epsilon\ (1{,}2159\ ; 1{,}6994)$$

Для построения доверительного интервала для коэффициента корреляции сделаем обратное преобразование Фишера:

$$r_{xy} \in \left(\frac{e^{2zl}-1}{e^{2zl}+1}\ ;\ \frac{e^{2zr}-1}{e^{2zr}+1}\right)$$

Для $\gamma = 0{,}95$:

$$r_{xy} \in (0{,}8545\ ; 0{,}9276)$$

Для $\gamma = 0{,}99$:

$$r_{xy} \in (0{,}8384\ ; 0{,}9353)$$

При увеличении уровня надежности получили более широкий доверительный интервал.

7. Осуществить проверку статистической гипотезы о равенстве коэффициента корреляции нулю при заданном уровне значимости $\alpha = 0,05$ сделать выводы.

Проверим гипотезу о равенстве нулю коэффициента корреляции. Вычислим $T_{\text{набл}}$ по формуле:

$$T_{\text{набл}} = \frac{\overline{r_{xy}}\sqrt{N-2}}{\sqrt{1 - r_{xy}^{2}}} = 0,8972\frac{10,7238}{0,4416} = 21,7876$$

Для уровня значимости $\alpha = 0,05$ и $k = N - 2 = 115$ было найдено $T_{\text{крит}} = 1,982$.

Исходя из того, что $T_{\text{набл}} > T_{\text{крит}}$, гипотеза о равенстве нулю коэффициента корреляции отвергается.

Выводы,

В ходе выполнения лабораторной работы были освоены основные понятия, связанные с корреляционной зависимостью между случайными величинами, статистическими гипотезами и проверкой их «справедливости».

ПРИЛОЖЕНИЕ А
ИСХОДНЫЙ КОД

```
import pandas as pd
import numpy as np
from lab2_y import inter_df as inter_df_y, means as means_y, df as df_y,
up_range as up_range_y, \
    low_range as low_Range_y, h as h_y, s as s_y, moments as m_y, k,
start_moment_1_emp as mean_y, nums as nums_y
from lab2 import inter_df as inter_df_x, means as means_x, df as df_x,
up_range as up_range_x, \
    low_range as low_Range_x, h as h_x, s as s_x, moments as m_x,
start_moment_1_emp as mean_x, nums as nums_x

df = pd.read_csv('sample.csv', header=None)

rows = []

for u_x, l_x in zip(up_range_x, low_Range_x):
    cols = []
    cond_x = (l_x <= df.iloc[:, 0]) & \
             (df.iloc[:, 0] < u_x) \
        if u_x < max(df.iloc[:, 0]) \
        else (l_x <= df.iloc[:, 0]) & (df.iloc[:, 0] <= u_x)

    rng = df.iloc[:, 1][cond_x]
    for u_y, l_y in zip(up_range_y, low_Range_y):
        cond_y = (l_y <= rng) & \
                 (rng < u_y) \
            if u_y < max(rng) \
            else (l_y <= rng) & (rng <= u_y)

        cols.append(sum(cond_y))

    rows.append(cols)

rows = np.array(rows)
cor_table = pd.DataFrame(data=rows, index=means_x, columns=means_y)
cor_table.to_csv('Двумерный интервальный ряд.csv')
C_x = means_x[int(len(means_x) / 2)]
C_y = means_y[int(len(means_y) / 2)]

v_x = (np.array(means_x) - C_x) / h_x
v_y = (np.array(means_y) - C_y) / h_y
```

```
cor_table.columns = v_x
cor_table = cor_table.set_index(v_y)

cor_table.to_csv('Корялиционная таблица.csv')
rows = []
for i in range(len(cor_table)):
    cols = []
    for j in range(len(cor_table.columns)):
        cols.append(cor_table.to_numpy()[i][j] * v_y[j] * v_x[i])
    rows.append(cols)

help_table = pd.DataFrame(data=rows, index=v_x, columns=v_y)
help_table.to_csv('Вспомогательная таблица.csv')
sums = help_table.to_numpy().sum(axis=0)
s = sums.sum()


cor = (s - len(df)*m_x[0]*m_y[0]) / (len(df) * (s_x/h_x)*(s_y/h_y))


zb = np.log((1+cor)/(1 - cor)) / 2
o_b = 1 / np.sqrt(len(df) - 3)
z_r = zb + 1.96 *o_b
z_l = zb - 1.96 *o_b

T_n = cor * np.sqrt(len(df) - 2) / np.sqrt(1 - cor**2)

if __name__=='__main__':
    print('Суммы столбцов:', sums)
    print('Сумма: ', s)
    print("Коэф корреляции: ", cor)
    print("S_v: ", (s_x/h_x))
    print("S_u: ", (s_y/h_y))
    print("v_b: ", m_x[0])
    print("u_b: ", m_y[0])
    print(
        '({}:{})'.format((np.exp(2 * z_l) - 1) / (np.exp(2 * z_l) + 1),
(np.exp(2 * z_r) - 1) / (np.exp(2 * z_r) + 1)))
    print('k: ', k - 2)
    print('T_krit: ', 1.943)
    print('T_n: ', T_n)

    print('Отвергаем' if T_n > 1.943 else 'Принимаем')
```