Лабораторные работы по математической статистике

Подготовили студенты группы ГЭ-13: Гришин Н.В.
Феоктистов М.С.

Исходные данные

X	Y	X	Y	X	Y	X	Y	X	Y
40	57,3	40,5	50,4	53,1	52,3	38	54,6	41,9	51,7
47,4	47,2	41,9	52,6	40,9	52	41,1	49,2	33,7	59,8
44,7	49,3	46,4	48,2	41,7	51,8	30	56,6	29	59,5
39,8	53	46,8	49,1	47,3	52,2	44,6	54,2	43,8	53,4
50,7	49,7	39,1	56,9	38	57	45,8	53,3	42,8	51,2
30,8	56,8	51	52	35,2	61,6	39,8	54,3	51	44,7
38,3	51,7	43,9	49,8	45,8	52,1	26,5	55,9	40,7	57
42,7	49,4	27	58,1	55,6	48,9	33	56,7	32,9	57,3
53,6	47,3	34,8	55,6	44,4	51,8	44,6	51,4	40,9	53,6
39,1	54,4	34,9	59	40,1	51	50,3	48,3	28,9	57,6
46,7	49,2	33,7	57,6	46,6	52,3	45,6	51,3	50,4	41,2
38,6	57,3	30,7	56,9	31,9	57,8	47,6	48,1	30,5	57
44,3	50,1	35,8	52,8	37,4	53,7	50,2	46,9	40,8	51,7
53,9	45,9	36,9	56,7	43	54,3	40,8	54	33,8	57,9
48,8	51	38,3	52,7	46,5	52,2	42,2	56,9	41	52,3
47	54,5	52,7	45,3	46	52,1	34,1	55,9	32,2	61,4
45	52	44,1	54	51,8	47,1	34,2	56,1	31,8	55,7
41,6	52,1	36,1	56,1	40,6	54,1	32,4	57,2	40,7	52,3
45,3	50,5	46,7	47,2	48,9	46,9	44,4	47,3	37,4	48
35	58,1	40,5	47,8	56,6	44,4	33,9	56	44,1	53,8

Лабораторная работа №1
Обработка статистических данных

Цель работы:

1. Изучить основные понятия выборочного метода.

2. Ознакомиться с методикой первичной обработки данных.

3. Получить эмпирические распределения измеримого признака, т.е. оценить распределение генеральной совокупности по сгруппированным данным.

ХОД РАБОТЫ:

1. Упорядочим исходные данные в порядке возрастания, получим следующую таблицу:

X	Y	X	Y	X	Y	X	Y	X	Y
26,5	41,2	34,8	49,2	40,5	52	43,9	54	46,8	56,9
27	44,4	34,9	49,3	40,5	52,1	44,1	54,1	47	56,9
28,9	44,7	35	49,4	40,6	52,1	44,1	54,2	47,3	57
29	45,3	35,2	49,7	40,7	52,1	44,3	54,3	47,4	57
30	45,9	35,8	49,8	40,7	52,2	44,4	54,3	47,6	57
30,5	46,9	36,1	50,1	40,8	52,2	44,4	54,4	48,8	57,2
30,7	46,9	36,9	50,4	40,8	52,3	44,6	54,5	48,9	57,3
30,8	47,1	37,4	50,5	40,9	52,3	44,6	54,6	50,2	57,3
31,8	47,2	37,4	51	40,9	52,3	44,7	55,6	50,3	57,3
31,9	47,2	38	51	41	52,3	45	55,7	50,4	57,6
32,2	47,3	38	51,2	41,1	52,6	45,3	55,9	50,7	57,6
32,4	47,3	38,3	51,3	41,6	52,7	45,6	55,9	51	57,8
32,9	47,8	38,3	51,4	41,7	52,8	45,8	56	51	57,9
33	48	38,6	51,7	41,9	53	45,8	56,1	51,8	58,1
33,7	48,1	39,1	51,7	41,9	53,3	46	56,1	52,7	58,1
33,7	48,2	39,1	51,7	42,2	53,4	46,4	56,6	53,1	59
33,8	48,3	39,8	51,8	42,7	53,6	46,5	56,7	53,6	59,5
33,9	48,9	39,8	51,8	42,8	53,7	46,6	56,7	53,9	59,8
34,1	49,1	40	52	43	53,8	46,7	56,8	55,6	61,4
34,2	49,2	40,1	52	43,8	54	46,7	56,9	56,6	61,6

x_{min}=26,5 x_{max}=56,6 y_{min}=41,2 y_{max}=61,6

2. Найдем размах варьирования:

$$R_x = \ x_{max} - x_{min} = 56{,}6 - 26{,}5 = 30{,}1$$

$$R_y = y_{max} - y_{min} = 20{,}4$$

3. Найдем число интервалов по формуле Стьюргесса: k = 1 + 3,2 lg n = 7 где n – объем выборки, k – целая часть полученного числа.

4. Найдем длину интервалов:

$$h_x = \frac{R_x}{k} = 4{,}3$$

$$h_y = 2{,}9$$

5. Найдем границы интервалов каждого признака таким образом, чтобы минимальное значение стало серединой первого интервала, а максимальное – серединой последнего.

$\alpha_0 = x_{min} - \frac{h_x}{2}$ $\beta_0 = y_{min} - \frac{h_y}{2}$

	24,35		39,7
	28,65		42,7
	32,95		45,6
	37,25		48,5
	41,55		51,4
	45,85		54,3
	50,15		57,2
	54,45		60,1
	58,75		63,1

6. Найдем середины получившихся интервалов:

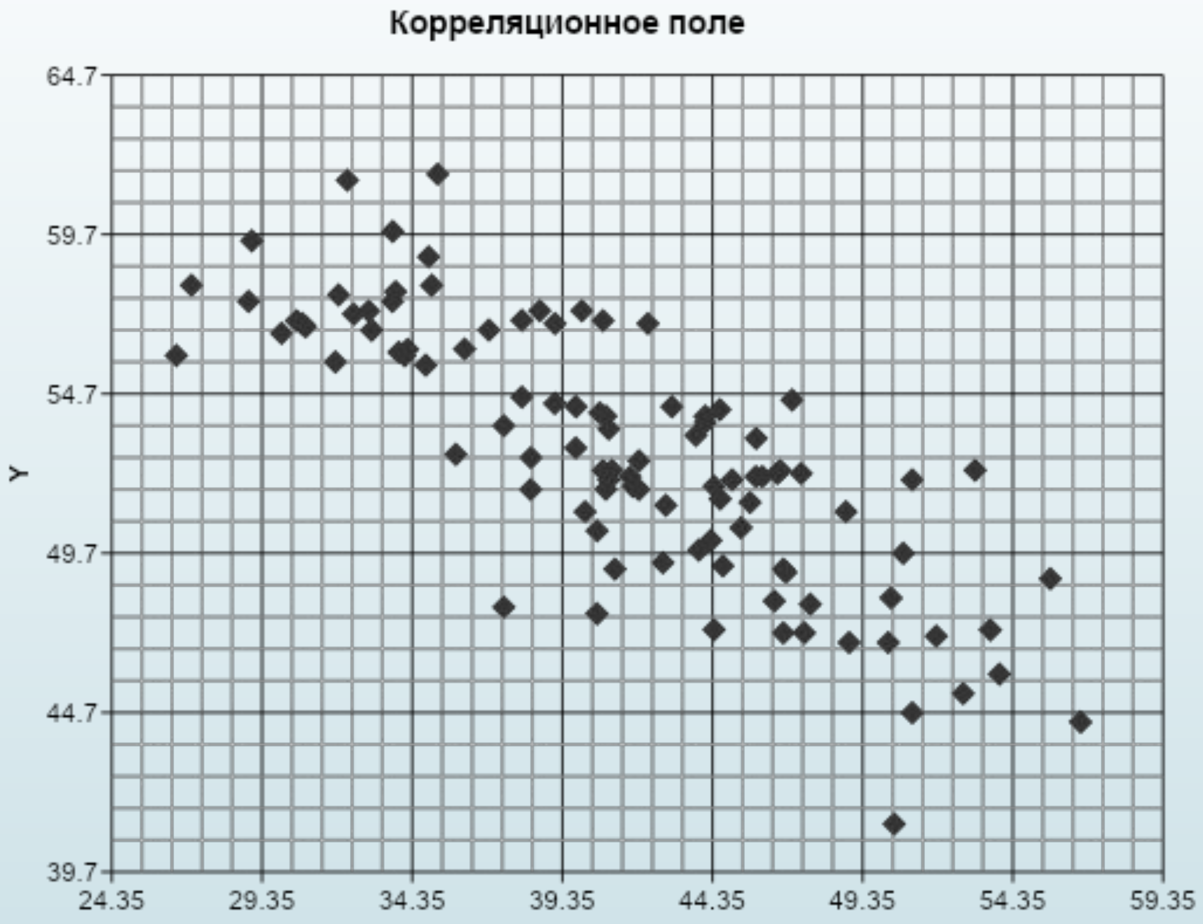
$x_1 = x_{min},$ $y_1 = y_{min},$

$x_2 = x_{min} + h_x$ $y_2 = y_{min} + h_y$

$x_3 = x_2 + h_x$ $y_3 = y_2 + h_y$

7. Построим корреляционное поле:

В двумерной системе координат оси OX и OY разбиваются на интервалы длиной h_x и h_y. За начало координат возьмите точку (α_0, β_0) т.е. начальные значения интервалов. Через границы интервалов α_i и β_i проводятся прямые, параллельные осям,- получают «сетку» из промежутков, в которую вносят (отмечают) исходные сто пар точек (x,y). Условимся, что при попадании какой-нибудь точки (x,y) на линию сетки, её относят к правому или верхнему квадрату.



8. Заполним корреляционную таблицу абсолютных частот

	0	0	0	0	0	0	1	0	1
	0	0	0	0	0	0	2	1	3
	0	0	0	2	1	5	5	0	13
	0	0	0	3	6	3	1	1	14
	0	0	1	11	13	4	2	0	31
	1	5	7	6	3	1	0	0	23
	1	5	5	2	0	0	0	0	13
	0	1	1	0	0	0	0	0	2
	2	11	14	24	23	13	11	2	100

В строках указываем середины x_i интервалов (α_{i-1}, α_i), в столбцах – середины y_i интервалов (β_{i-1}, β_i), а в соответствующую ячейку таблицы записываем число точек n_{ij}, попавших в аналогичную
(i, j) -ую ячейку сетки корреляционного поля.

9. Заполним таблицу «Статистическая совокупность» для каждого признака в отдельности:

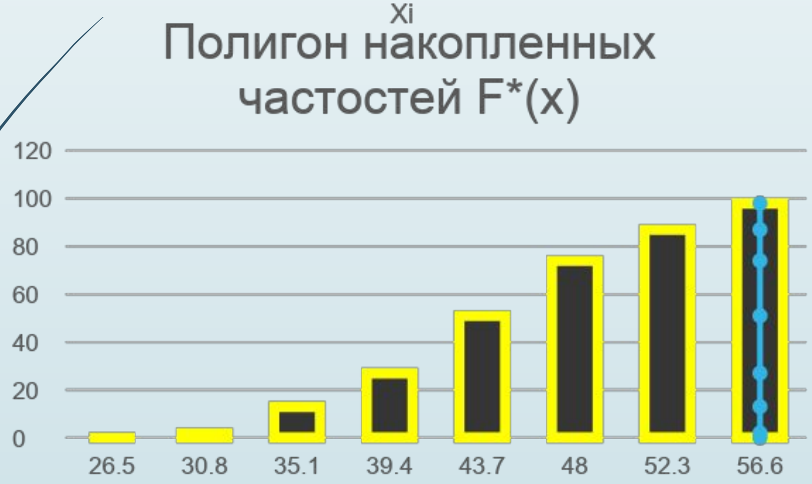
Статистическая совокупность измеримого признака X

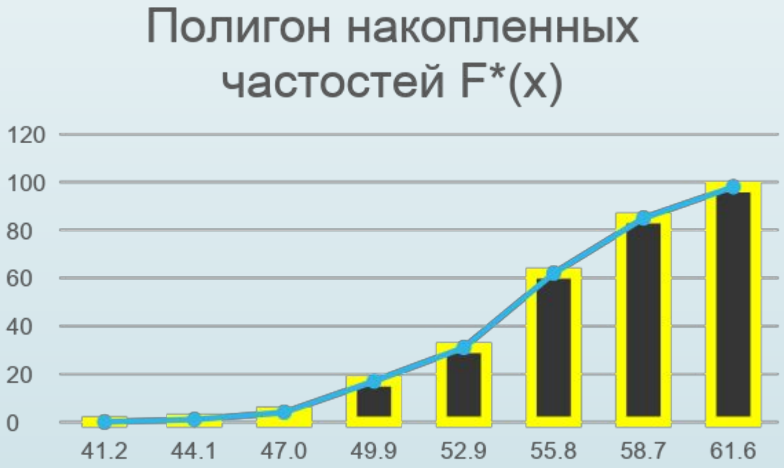
24,35-28,65	26,5	2	0,02	0	0	0,0047
28,65-32,95	30,8	11	0,11	2	0,02	0,0256
32,95-37,25	35,1	14	0,14	13	0,13	0,0326
37,25-41,55	39,4	24	0,24	27	0,27	0,0558
41,55-45,85	43,7	23	0,23	51	0,51	0,0535
45,85-50,15	48	13	0,13	74	0,74	0,0302
50,15-54,45	52,3	11	0,11	87	0,87	0,0256
54,45-58,75	56,6	2	0,02	98	0,98	0,0047

Статистическая совокупность измеримого признака Y

39,7-42,7	41,2	1	0,01	0	0	0,0034
42,7-45,6	44,1	3	0,03	1	0,01	0,0103
45,6-48,5	47,0	13	0,13	4	0,04	0,0446
48,5-51,4	49,9	14	0,14	17	0,17	0,0480
51,4-54,3	52,9	31	0,31	31	0,31	0,1064
54,3-57,2	55,8	23	0,23	62	0,62	0,0789
57,2-60,1	58,7	13	0,13	85	0,85	0,0446
60,1-63,1	61,6	2	0,02	98	0,98	0,0069

10. Построим полигон и гистограмму распределения, затем – полигон накопленных частостей:





Вывод: в данной лабораторной работе мы изучили основные понятия выборочного метода, ознакомились с методикой первичной обработки данных, и получили эмпирические распределения измеримого признака, т.е. оценили распределение генеральной совокупности по сгруппированным данным.

Лабораторная работа №2
Статистические точечные оценки генеральных параметров.

Цель работы:

Оценить генеральные параметры по сгруппированным данным.

Мода – значение в множестве наблюдений, которое встречается наиболее часто.

Медиана – возможное значение признака, которое делит ранжированную совокупность (вариационный ряд выборки) на две части.

Асимметрия представляет собой числовое отображение степени отклонения графика распределения показателей от симметричного графика распределения. Если асимметрия больше нуля, то она положительная и левосторонняя. Если меньше нуля, то она отрицательна и правосторонняя.

Эксцесс – показатель остроты пика графика распределения. Эксцесс симметричного распределения равен нулю. Если эксцесс больше нуля, то график плосковершинный, если меньше – островершинный.

ХОД РАБОТЫ

1) **Заполним расчетную таблицу:**

Таблица 1

Расчет выборочных оценок признака X

26,5	2	0,02	0,5	-15,0	4,5	-67,0	1002,8
30,8	11	0,11	3,4	-10,7	12,5	-133,4	1422,6
35,1	14	0,14	4,9	-6,4	5,7	-36,1	229,6
39,4	24	0,24	9,5	-2,1	1,0	-2,1	4,4
43,7	23	0,23	10,1	2,2	1,1	2,6	5,7
48	13	0,13	6,2	6,5	5,6	36,3	237,2
52,3	11	0,11	5,8	10,8	12,9	140,0	1516,6
56,6	2	0,02	1,1	15,1	4,6	69,4	1049,7
Σ	100	1	41,464		47,88	9,57	5468,7

Аналогичная таблица заполняется для измеримого признака У:

Таблица 2

Расчет выборочных оценок признака Y

41,2	1	0,01	0,41	-11,72	1,37	-16,08	188,38
44,1	3	0,03	1,32	-8,80	2,32	-20,45	180,00
47,0	13	0,13	6,11	-5,89	4,51	-26,52	156,13
49,9	14	0,14	6,99	-2,97	1,24	-3,68	10,93
52,9	31	0,31	16,39	-0,06	0,00	0,00	0,00
55,8	23	0,23	12,83	2,86	1,88	5,36	15,30
58,7	13	0,13	7,63	5,77	4,33	24,98	144,12
61,6	2	0,02	1,23	8,68	1,51	13,10	113,77
Σ	100	1	52,92		17,15	-23,30	808,6

2) **Найдем выборочные оценки для признака X находим по данным таблицы 8 и формулам для сгруппированных данных:**

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\, n_i = \sum_{i=1}^{r} x_i\, p_i^* = 41{,}464$$

$$D_B = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{r} (x_i - \bar{x})^2\, p_i^* = 47{,}88$$

$$\sigma_B = \sqrt{D_B} = 7{,}34$$

$$A_B = \frac{\mu_3}{(\mu_2)^{3/2}} = 0{,}029$$

$$E_B = \frac{\mu_4}{(\mu_2)^2} - 3 = -0{,}61$$

Аналогичные выборочные оценки для признака У:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\, n_i = \sum_{i=1}^{r} y_i\, p_i^* = 52{,}92$$

$$D_B = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{r} (y_i - \bar{y})^2\, p_i^* = 17{,}15$$

$$\sigma_B = \sqrt{D_B} = 4{,}14$$

$$A_B = \frac{\mu_3}{(\mu_2)^{3/2}} = -0{,}33$$

$$E_B = \frac{\mu_4}{(\mu_2)^2} - 3 = -0{,}25$$

3) Найдем исправленные оценки признака Х:

- выборочное среднее $\bar{x} = 41{,}46$
- исправленная дисперсия $S^2 = 48{,}37$
- исправленное среднеквадратичное отклонение S=$\sqrt{S^2}$ =6,95
- исправленная асимметрия $A^* = 0{,}03$
- исправленный эксцесс $E^* = -0{,}58$

Исправленные оценки признака У:

- выборочное среднее $\bar{y} = 52.92$
- исправленная дисперсия $S^2 = 17.32$
- исправленное среднеквадратичное отклонение S=$\sqrt{S^2}$ =4.16
- исправленная асимметрия $A^* = -0.33$
- исправленный эксцесс $E^* = -0.20$

4) **Найдем моду и медиану по сгруппированным данным для признака X:**

$$M_o^* = 41{,}16$$
$$M_i^* = 41{,}37$$

Для признака У:

$$M_o^* = 53{,}37$$
$$M_i^* = 53{,}18$$

5) **Коэффициент вариации, т.е. среднее квадратическое отклонение S в процентах от среднего значения ряда измерений для признака X:**

$$v = \frac{S}{\bar{x}} \cdot 100\% = 16{,}8\%$$

Для признака У:

$$v = \frac{S}{\bar{y}} \cdot 100\% = 7{,}86\%$$

Выводы (для признака X):

а) $\bar{x} = M_o^* = M_i^*$

б) $A^* = 0{,}03$ – больше нуля, значит полигон распределения скошен, правая ветвь длиннее левой, начиная от вершины: левосторонняя асимметрия. А* близко к нулю. $E^* = -0{,}58$ – меньше нуля, гистограмма – плосковершинная (по сравнению с нормальным распределением).

в) Коэффициент вариации равен 16,8%.

(для признака У):

а) $\bar{y} = M_o^* = M_i^*$

б) $A^* = -0{,}33$ – меньше нуля, левая ветвь длинее правой, начиная от вершины: правосторонняя асиметрия. А* близко к нулю. $E^* = -0{,}20$ – меньше нуля, гистограмма – плосковершинная (по сравнению с нормальным распределением).

в) Коэффициент вариации равен 7,86%

Лабораторная работа №3
Статистическая проверка статистической гипотезы о совпадении с нормальным распределением одного измеримого признака генеральной совокупности

ЦЕЛЬ РАБОТЫ: Ознакомиться с основными задачами статистической проверки гипотез, с часто используемыми методами проверки гипотезы нормальности распределения. Изучить решение задачи о согласованности теоретического и статистического распределений.

Ход работы:

1. Для признака X.

Выясним, чему равен коэффициент вариации в лабораторной работе №2, проверка нормальности распределения проводится только при выполнении условия $v < 33\%$.

$$v = \frac{S}{\bar{x}} \cdot 100\% = 16{,}8$$

Проверку нормальности распределения измеримого признака

X проводить имеет смысл, т.к. $v < 33\%$.

2. Выпишем статистики распределения измеримого признака, например X, из лабораторной работы № 2: $\bar{x}$ и S_x

$$\bar{x} = 41{,}46\ ;\ S_x = 6{,}95$$

3. Формулируем статистическую гипотезу Но: генеральная совокупность измеримого признака X, из которой извлечена выборка, распределена по нормальному закону при данном уровне значимости α =0,05, с плотностью

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/2\sigma^2},\ \text{где}$$

a и σ – параметры нормального распределения.

4. Выполним проверку гипотезы Но по критерию Пирсона.

Статистика для проверки: $X^2 = \sum_{i=1}^{r} \frac{(n_i - n_i')^2}{n_i'}$

Случайная величина, распределенная по закону «хи-квадрат», с к- степенями свободы: к = r - 3, где r – число интервалов;

n_i - наблюдаемая абсолютная частота, соответствующая i-тому

интервалу; n_i' - теоретическая частота: $n_i' = n \cdot p_i'$;

$p_i' = \Phi\left(\frac{a_i - \overline{x}}{S_x}\right) - \Phi\left(\frac{a_{i-1} - \overline{x}}{S_x}\right) = \Delta\Phi$ – теоретическая вероятность попадания случайной величины X, распределенной нормально, в интервал a_{i-1} - a_i (значение функции Лапласа можно найти по Таблице)

Итак, для осуществления проверки по критерию Пирсона, необходимо:

- объединить интервалы (смотри лаб. работу № 2) с абсолютными частотами n_i, меньшими 5 , суммируя частоты;

- отметить, чему равно теперь r - число интервалов;
- записать число к - степеней свободы и по таблицам найти $X^2_{\text{кр}}(\alpha, k)$
- заполнить расчетную таблицу для вычисления $X^2 = \sum_{i=1}^{r} \frac{(n_i - n_i')^2}{n_i'}$

Выпишем из лабораторных работ № 1 и № 2 границы интервалов и абсолютные частоты в них

24,35-28,65	26,5	2
28,65-32,95	30,8	11
32,95-37,25	35,1	14
37,25-41,55	39,4	24
41,55-45,85	43,7	23
45,85-50,15	48	13
50,15-54,45	52,3	11
54,45-58,75	56,6	2

Видим, что в первом и последнем интервалах абсолютная частота меньше пяти. Объединяем первые два и последние два интервала, число интервалов r равно теперь 6, значит число степеней свободы к = r – 3 = 3 и $X^2_{кр}(\alpha, k) = X^2_{кр}(0{,}05; 3) = 7{,}8$.

5. Заполняем расчетную таблицу:

24,35	32,95	13	-2,46	-0,4931	0,1048	11	0,36
32,95	37,25	14	-1,22	-0,3883	0,1592	16	0,23
37,25	41,55	24	-0,61	-0,2291	0,2331	23	0,02
41,55	45,85	23	0,01	0,004	0,2184	22	0,06
45,58	50,15	13	0,59	0,2224	0,172	17	1,03
50,15	58,75	13	1,25	0,3944	0,0992	10	0,96
			2,49	0,4936			
		100	0,05	0,0039	0,9867	100	

6.Найдем интервальные оценки α и σ нормально распределенной генеральной совокупности X:

$$t_v = t(n, v) \;\; v = (1 - a) \text{ и } q = q(v, n)$$

$t_v = t(100; 0{,}95) = 1{,}984$ и $q = q(100; 0{,}95) = 0{,}143$; тогда

$$\overline{x} - \frac{t_v \cdot s_x}{\sqrt{n}} < a < \overline{x} + \frac{t_v \cdot s_x}{\sqrt{n}}$$

$$40{,}08 < a < 42{,}84$$

$$S_x(1 - q) < \sigma < S_x(1 + q)$$

$$5{,}96 < \sigma < 7{,}94$$

(Вывод) Получили, что $X^2_{набл} = 2{,}66$ меньше, чем $X^2_{кр}(\alpha, k) = 7{,}8$, значит гипотеза нормальности распределения принимается. Построим график плотности теоретического распределения f(x) и сравним его с полигоном относительных частот. Принимаем a =41,46, σ =6,95, получили теоретическую функцию распределения измеримого признака X

$$f(x) = \frac{1}{6{,}95\sqrt{2\pi}} e^{-(x-41{,}46)^2/2\cdot 6{,}95^2}$$

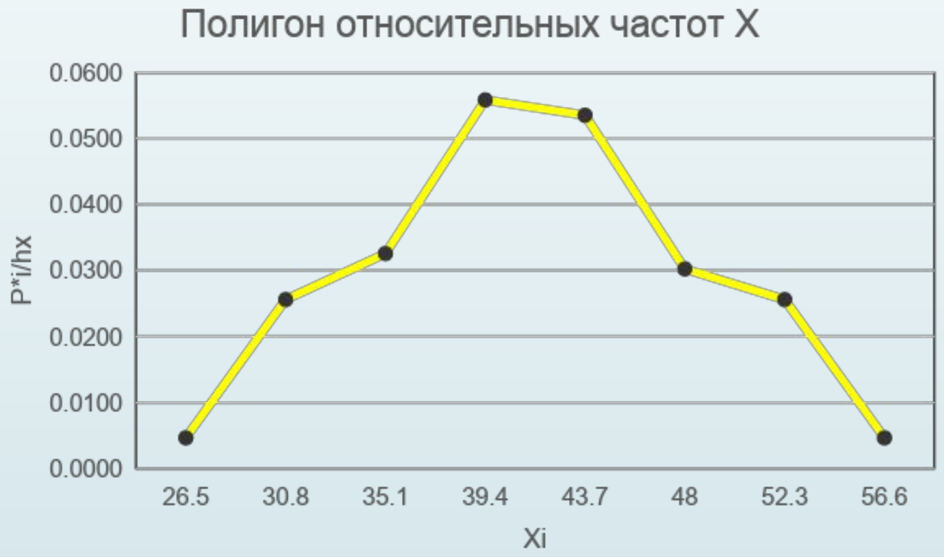


График несколько отличается от полигона относительных частот эмпирически полученной в лабораторной работе функции $f_i^* = \frac{P_i^*}{h_x}$, но отражает главные свойства этой функции – её интервалы монотонности, экстремум; то есть можно сказать, что теоретически полученная функция распределения хорошо согласуется с эмпирической функцией распределения.

II. Аналогично для признака У:

1. Так как коэффициент вариации $v = \frac{S}{\bar{x}} \cdot 100\% = 7{,}86$ и $v < 33\%$,

проверку нормальности распределения измеримого признака

Y проводить имеет смысл.

2. $\bar{y} = 52{,}92$ и $S_y = 4{,}16$

3. Формулируем статистическую гипотезу Но: генеральная совокупность измеримого признака Y, из которой извлечена выборка, распределена по нормальному закону при данном уровне значимости α =0,05, с плотностью

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/2\sigma^2}$$

4. Выпишем из лабораторных работ № 1 и № 2 границы интервалов и абсолютные частоты в них

39,7-42,7	41,2	1
42,7-45,6	44,1	3
45,6-48,5	47,0	13
48,5-51,4	49,9	14
51,4-54,3	52,9	31
54,3-57,2	55,8	23
57,2-60,1	58,7	13
60,1-63,1	61,6	2

Видим, что в первом, втором и последнем интервалах абсолютная частота меньше пяти. Объединяем первые два и последние два интервала, число интервалов r равно теперь 5, значит число степеней свободы k = 2 и $X^2_{\text{кр}}(\alpha, k) = X^2_{\text{кр}}(0{,}05; 2) = 6.$

5.

39,7	48,5	17	-3,18	-0,4992	0,1438	15	0,27
48,5	51,4	14	-1,06	-0,3554	0,2111	21	2,39
51,4	54,3	31	-0,37	-0,1443	0,2736	27	0,48
54,3	57,2	23	0,33	0,1293	0,2192	22	0,05
57,2	63,1	15	1,03	0,3485	0,1444	14	0,02
			2,45	0,4929			
		100	-0,80	-0,0282	0,99	100	3,22

6. Найдем интервальные оценки α и σ нормально распределенной генеральной совокупности Y:

$t_v = t(100; 0{,}95) = 1{,}984$ и $q = q(100; 0{,}95) = 0{,}143$; тогда

$$52{,}09 < a < 53{,}75$$

$$3{,}57 < \sigma < 4{,}75$$

7. (Вывод). Получили, что $X^2_{\text{набл}} = 2{,}66$ меньше, чем $X^2_{\text{кр}}(\alpha, k) = 7{,}8$, значит гипотеза нормальности распределения принимается. Построим график плотности теоретического распределения f(x) и сравним его с полигоном относительных частот: принимаем a =52,92, σ =4,16, получили теоретическую функцию распределения измеримого признака Y

$$f(\text{y}) = \frac{1}{4{,}16\sqrt{2\pi}} e^{-(\text{y}-52{,}92)^2/34{,}61}$$

График несколько отличается от полигона относительных частот эмпирически полученной в лабораторной работе функции $f_i^* = \frac{P_i^*}{\ldots}$, но отражает главные свойства этой функции – её интервалы

Лабораторная работа №4
Корреляционная зависимость между двумя измеримыми признаками.
Расчет коэффициентов уравнения линейной регрессии, их статистическая оценка

ЦЕЛЬ РАБОТЫ: Ознакомиться с основными понятиями и методами исследования корреляционной зависимости на примере линейной корреляции. Сделать статистическое оценивание коэффициентов регрессии. Уровень значимости принять равным 0,05.

ХОД РАБОТЫ

1. Выпишем результаты лабораторной работы № 2:

$$\bar{x} = 41{,}464$$

$$\bar{y} = 52{,}92$$

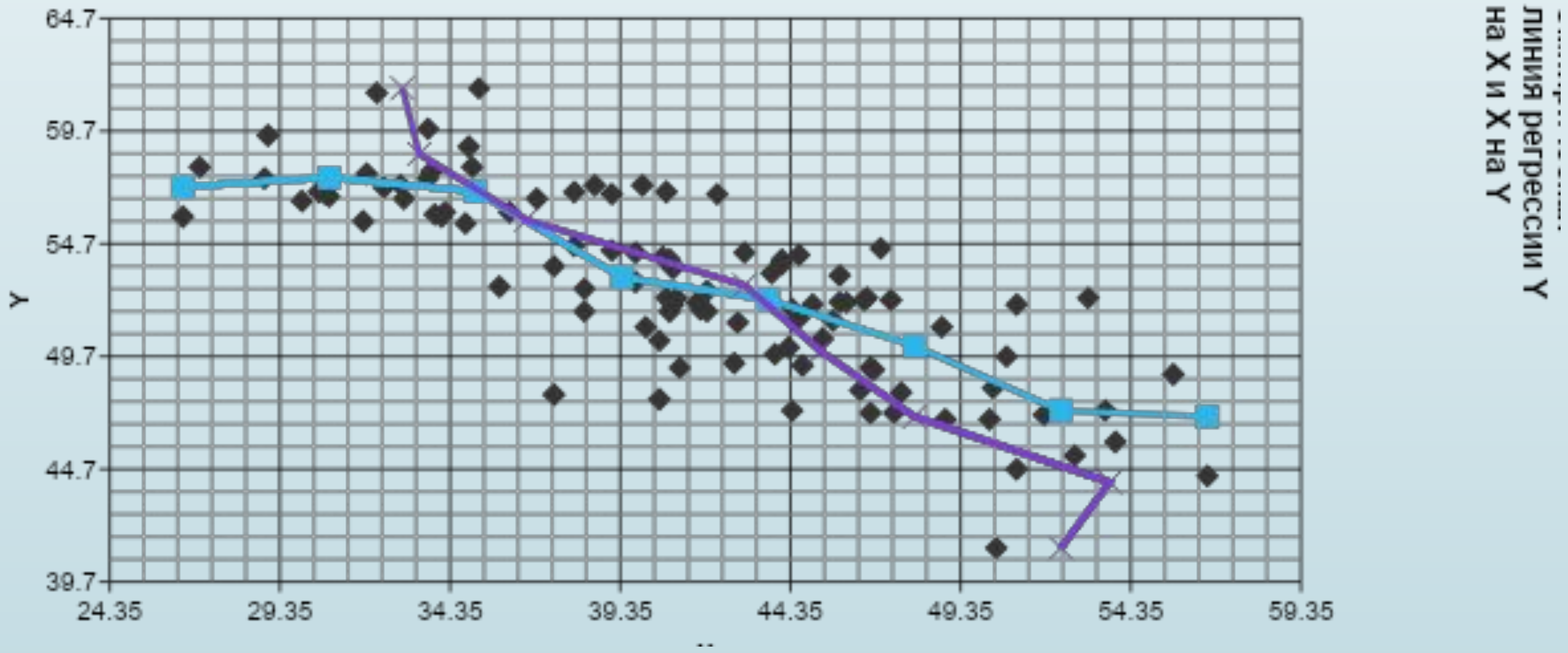
2. 2. Запишем корреляционную таблицу и вычислим условные средние $\overline{x}_y$ и $\overline{y}_x$, которые вычисляются по формулам:

$$\overline{y}_{x=x_i} = \frac{1}{n_{x_i}} \sum_{j=1}^{r} y_j \cdot n_{x_i y_j}$$

$$\overline{x}_{y=y_j} = \frac{1}{n_{y_j}} \sum_{i=1}^{r} x_i \cdot n_{x_i y_j}$$

x y	26,5	30,8	35,1	39,4	43,7	48	52,3	56,6	ny	
41,2	0	0	0	0	0	0	1	0	1	52,30
44,1	0	0	0	0	0	0	2	1	3	53,73
47,0	0	0	0	2	1	5	5	0	13	48,00
49,9	0	0	0	3	6	3	1	1	14	45,24
52,9	0	0	1	11	13	4	2	0	31	43,01
55,8	1	5	7	6	3	1	0	0	23	36,60
58,7	1	5	5	2	0	0	0	0	13	33,45
61,6	0	1	1	0	0	0	0	0	2	32,95
nx	2	11	14	24	23	13	11	2	100	
	57,23	57,63	57,02	53,22	52,22	50,17	47,29	47,03		

3. В корреляционном поле построим эмпирические линии регрессии Y на X и X на Y



4. Аппроксимируем эти ломанные прямой $y = ax + b$ (используя метод наименьших квадратов), тем самым подберем графики функции регрессии.

Суть метода наименьших квадратов.

Задача заключается в нахождении коэффициентов линейной зависимости, при которых функция 2-х переменных a и b будет равна:

$$F(a;b) = \sum_{i=1}^{\infty}\left(y_i - (ax_i + b)\right)^2$$

Найдем минимальное значение этой функции.

$$\begin{cases} \dfrac{\partial F}{da} = -2\sum_{i=1}^{8}(y_i - (ax_i + b) \cdot x_i \\ \dfrac{\partial F}{db} = -2\sum_{i=1}^{8}(y_i - (ax_i + b) \end{cases} \begin{cases} \dfrac{\partial F}{da} = 0 \\ \dfrac{\partial F}{db} = 0 \end{cases}$$

$$\begin{cases} -2\sum_{i=1}^{8}(y_i - (ax_i + b) \cdot x_i = 0 \\ -2\sum_{i=1}^{8}(y_i - (ax_i + b) = 0 \end{cases} \qquad \begin{cases} \sum_{i=1}^{8} y_i \cdot x_i - a\sum_{i=1}^{8} x_i^2 - b\sum_{i=1}^{8} x_i = 0 \\ \sum_{i=1}^{8} y_i - a\sum_{i=1}^{8} x_i - 8b = 0 \end{cases}$$

Подставляем все значения и находим коэффициенты a и b

4.1 (для У на Х)

Выпишем в таблицу средние значения x_i и условные средние $\overline{y}_x$.

	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	∑
	26,5	30,8	35,1	39,4	43,7	48	52,3	56,6	332,4
	57,2	57,6	57,0	53,2	52,2	50,2	47,3	47,0	421,8
	1516,557	1774,88	2001,416	2096,924	2282,171	2408,018	2473,45	2661,817	17215,23
	702,25	948,64	1232,01	1552,36	1909,69	2304	2735,29	3203,56	14587,8

$$\begin{cases} \sum_{i=1}^{8} y_i \cdot x_i - a\sum_{i=1}^{8} x_i^2 - b\sum_{i=1}^{8} x_i = 0 \\ \sum_{i=1}^{8} y_i - a\sum_{i=1}^{8} x_i - 8b = 0 \end{cases}$$

Подставляем все значения в данную систему и находим коэффициенты a и b.

$$\begin{cases} a = -0{,}39991 \\ b = 69{,}34115 \end{cases}$$

Линейное уравнение y=ax+b принимает следующий вид:

y=−0,39991x +69,34115

Для этой функции найдем значения $\hat{y}_i$, и подставим в таблицу. Вычислим сумму квадратов отклонений исходных данных y_i от теоретически рассчитанных $\hat{y}_i$ и занесем в таблицу.

	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	∑
	26,5	30,8	35,1	39,4	43,7	48	52,3	56,6	332,4
	57,2	57,6	57,0	53,2	52,2	50,2	47,3	47,0	421,8
	1516,557	1774,88	2001,416	2096,924	2282,171	2408,018	2473,45	2661,817	17215,23
	702,25	948,64	1232,01	1552,36	1909,69	2304	2735,29	3203,56	14587,8
	58,74	57,02	55,30	53,58	51,87	50,15	48,43	46,71	
	-1,52	0,60	1,72	-0,36	0,36	0,02	-1,13	0,32	
	2,30	0,36	2,94	0,13	0,13	0,00	1,28	0,10	7,25

4.2 (для X на Y)

Выпишем в таблицу средние значения y_i и условные средние $\overline{x}_y$.

	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	∑
	52,3	53,73333	48	45,23571	43,00645	36,59565	33,44615	32,95	345,2673
	41,2	44,1	47,0	49,9	52,9	55,8	58,7	61,6	411,2
	2154,76	2370,408	2257,371	2259,201	2273,198	2040,992	1962,811	2029,72	17348,46
	2735,29	2887,271	2304	2046,27	1849,555	1339,242	1118,645	1085,703	15365,98

$$\begin{cases} \sum_{i=1}^{8} y_i \cdot x_i - a\sum_{i=1}^{8} x_i^2 - b\sum_{i=1}^{8} x_i = 0 \\ \sum_{i=1}^{8} y_i - a\sum_{i=1}^{8} x_i - 8b = 0 \end{cases}$$

Подставляем все значения в данную систему и находим коэффициенты a и b.

$$\begin{cases} a = -0.85692 \\ b = 88.38319 \end{cases}$$

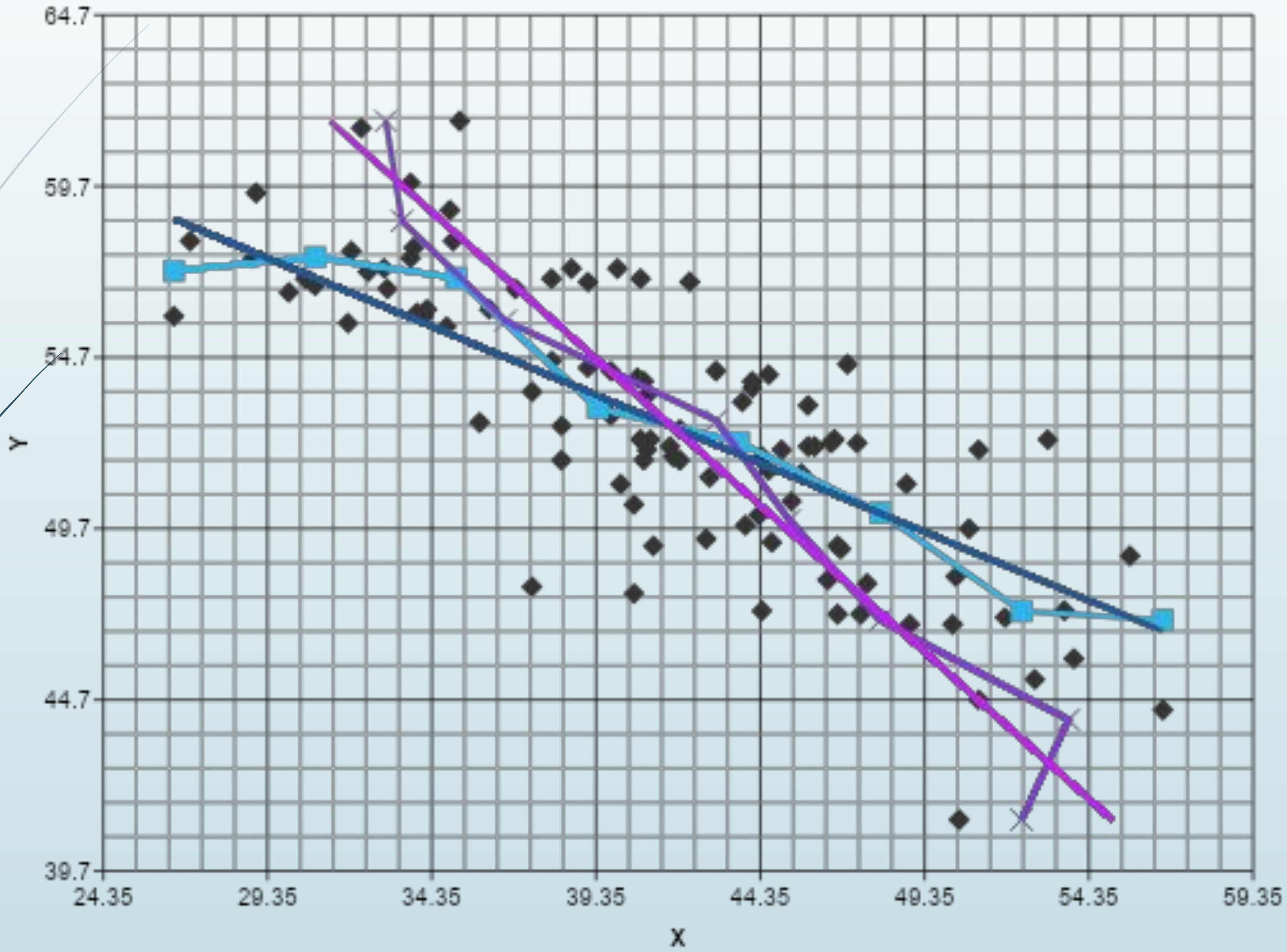
Линейное уравнение y=ax+b принимает следующий вид:

y=−0.85692x + 88.38319

Для этой функции найдем значения $\hat{x}_i$, и подставим в таблицу. Вычислим сумму квадратов отклонений исходных данных x_i от теоретически рассчитанных $\hat{x}_i$ и занесем в таблицу.

	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	∑
	52,3	53,73333	48	45,23571	43,00645	36,59565	33,44615	32,95	345,2673
	41,2	44,1	47,0	49,9	52,9	55,8	58,7	61,6	411,2
	2154,76	2370,408	2257,371	2259,201	2273,198	2040,992	1962,811	2029,72	17348,46
	2735,29	2887,271	2304	2046,27	1849,555	1339,242	1118,645	1085,703	15365,98
	58,74	57,02	55,30	53,58	51,87	50,15	48,43	46,71	
	-1,52	0,60	1,72	-0,36	0,36	0,02	-1,13	0,32	
	2,30	0,36	2,94	0,13	0,13	0,00	1,28	0,10	7,25

5. На корреляционном поле достроим графики функции регрессии.



6. Чтобы установить оценку тесноты корреляционной зависимости необходимо оценить коэффициент корреляции. В случае линейной зависимости коэффициент корреляции определим как:

$$r = \pm\sqrt{a_{yx} \cdot a_{xy}} = \pm\sqrt{-0{,}39991 \cdot (-0.85692)} = 0{,}59$$

7. $tg(\varphi) = \left|\frac{1-a_{yx}\cdot a_{xy}}{a_{yx}+a_{xy}}\right| = 0{,}52$

$\varphi = 28^{\circ}$

Вывод: В данной л.р. мы ознакомились с основными понятиями и методами исследования корреляционной зависимости на примере линейной корреляции. Сделали статистическое оценивание коэффициентов регрессии. Мы получили, что коэффициент корреляции оказался равным 0,52.