

Министерство образования и науки Российской Федерации
Уральский федеральный университет
имени первого Президента России Б.Н. Ельцина

Чердынцева Г.А., Кравченко Н.М., Трясцина Т.С.

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Методические указания и варианты к расчетно-графической работе
для студентов физических специальностей ФТИ

Екатеринбург

УрФУ

2014

Лабораторная работа №1

Первичная обработка результатов наблюдений двух измеримых признаков и статистические точечные оценки генеральных параметров.

Цель работы. Ознакомиться с методикой первичной обработки двумерной выборки, составить интервальные и дискретные вариационные ряды распределения признаков X и Y , эмпирическую функцию распределения одного из признаков (X), построить полигон, гистограмму и график эмпирической функции распределения признака X . Оценить генеральные параметры признака X .

Порядок выполнения работы.

1. Найти минимальные (x_{\min}, y_{\min}) и максимальные (x_{\max}, y_{\max}) значения выборочной совокупности X и Y . Установить границы интервалов α_i для признака $X: \alpha_i = \alpha_{i-1} + h_x$, $i = \overline{1, l}$, где α_0 – начало первого интервала, h_x – длина интервалов, l – число интервалов ($\alpha_0 < x_{\min} < \alpha_1 < \dots < x_{\max} < \alpha_l$). Для признака $Y: \beta_j = \beta_{j-1} + h_y$, $j = \overline{1, m}$, где β_0 – начало первого интервала, h_y – длина интервалов, m – число интервалов ($\beta_0 < y_{\min} < \beta_1 < \dots < y_{\max} < \beta_m$).

2. По данным выборки заполнить корреляционную таблицу абсолютных частот.

3. Найти середины интервалов x_i, y_j и составить вариационный ряд каждого признака.

Вычислить относительные частоты: $n_{\text{отн}}(x) = \frac{n_{x_i}}{n}$, $n_{\text{отн}}(y) = \frac{n_{y_j}}{n}$, накопленные частоты:

$$n_{\text{накоп}}(x) = \sum_{x_i < x} \frac{n_{x_i}}{n}, \quad n_{\text{накоп}}(y) = \sum_{y_j < y} \frac{n_{y_j}}{n} \quad \text{и плотности относительных частот} \quad \frac{n_{\text{отн}}(x)}{n_x}, \quad \frac{n_{\text{отн}}(y)}{n_y}.$$

4. Построить полигон частот и гистограмму относительных частот. Найти эмпирическую функцию распределения признака X по формуле $F^*(x) = \sum_{x_i < x} \frac{n_{x_i}}{n}$.

5. Найти выборочное среднее, выборочную дисперсию, а также несмещенные оценки дисперсии и среднеквадратического отклонения генеральной совокупности признаков X и Y по формулам:

$$\begin{aligned} \bar{x}_g &= \sum_{i=1}^l x_i \cdot \frac{n_{x_i}}{n}, & D_{xg} &= \sum_{i=1}^l (x_i - \bar{x}_g)^2 \frac{n_{x_i}}{n}, \\ \bar{y}_g &= \sum_{j=1}^m y_j \cdot \frac{n_{y_j}}{n}, & D_{yg} &= \sum_{j=1}^m (y_j - \bar{y}_g)^2 \frac{n_{y_j}}{n}, \\ S_x^2 &= \frac{n}{n-1} D_{xg}, & s_x &= \sqrt{S_x^2}, \\ S_y^2 &= \frac{n}{n-1} D_{yg}, & s_y &= \sqrt{S_y^2}. \end{aligned}$$

Рассмотрим на примере данной выборки выполнение первичной обработки данных и оценим параметры генеральной совокупности. Система статистических признаков (X, Y) задана таблицей. Объем выборки, начало первого интервала и рекомендуемая длина интервала для каждого признака указаны.

Таблица 1

X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
73	577	63	495	74	578	70	558	79	627	63	496
69	548	80	635	68	537	61	480	77	607	63	495
72	575	71	564	69	550	62	486	78	618	67	530
72	573	74	583	71	559	63	500	66	521	68	534
65	519	68	534	60	479	71	560	63	497	55	437
67	530	65	518	56	446	65	517	69	546	56	446
56	443	73	574	71	565	70	558	74	582	58	460
70	555	57	454	68	538	70	559	68	538	70	551
63	502	71	566	66	520	63	496	62	495	59	465
64	506	66	522	60	478	73	580	70	554	68	543
70	554	76	599	70	554	68	542	70	552	69	550
67	535	70	554	69	542	59	462	65	511	63	497
60	478	68	542	72	566	64	504	70	555	70	558

Объем выборки $n = 78$ X Y

Начало первого интервала 53 420

Длина интервала 5 34

Результаты обработки данных.

1. Для данной выборки минимальные и максимальные значения X и Y равны:

$$x_{\min} = 55, \quad x_{\max} = 80,$$

$$y_{\min} = 437, \quad y_{\max} = 635.$$

Найдем границы интервалов α_i и β_j :

$$\alpha_0 = 53, \quad \alpha_1 = 58, \quad \alpha_2 = 63, \quad \alpha_3 = 68, \quad \alpha_4 = 73, \quad \alpha_5 = 78; \quad \alpha_6 = 80;$$

$$\beta_0 = 420, \quad \beta_1 = 454, \quad \beta_2 = 488, \quad \beta_3 = 522, \quad \beta_4 = 556, \quad \beta_5 = 590, \quad \beta_6 = 624, \quad \beta_7 = 658.$$

2. Составим корреляционную таблицу абсолютных частот.

Таблица 2

$Y \backslash X$	[53, 58)	[58, 63)	[63, 68)	[68, 73)	[73, 78)	[78, 83)	n_{y_j}
[420, 454)	4						4
[454, 488)	1	8					9
[488, 522)		1	16				17
[522, 556)			4	21			25
[556, 590)				12	6		18
[590, 624)					2	1	3
[624, 658)						2	2
n_{x_i}	5	9	20	33	8	3	78

3. Вычислим относительные, накопленные частоты, плотность относительных частот. Составим дискретные и интервальные вариационные ряды для признаков X и Y . Полученные данные занесем в таблицы (табл.3, табл.4).

Таблица 3

№	$[\alpha_{i-1}, \alpha_i)$	x_i	n_{x_i}	$\frac{n_{x_i}}{n}$	$n_{\text{накоп}}$	$\frac{n_{\text{отн}}}{h_x}$
1	[53, 58)	55,5	5	0,064	0	0.013
2	[58, 63)	60,5	9	0,115	0,064	0,023
3	[63, 68)	65,5	20	0,257	0,179	0,051
4	[68, 73)	70,5	33	0,423	0,436	0,085
5	[73, 78)	75,5	8	0,103	0,850	0,021
6	[78, 83)	80,5	3	0,038	0,962 1,000	0,008

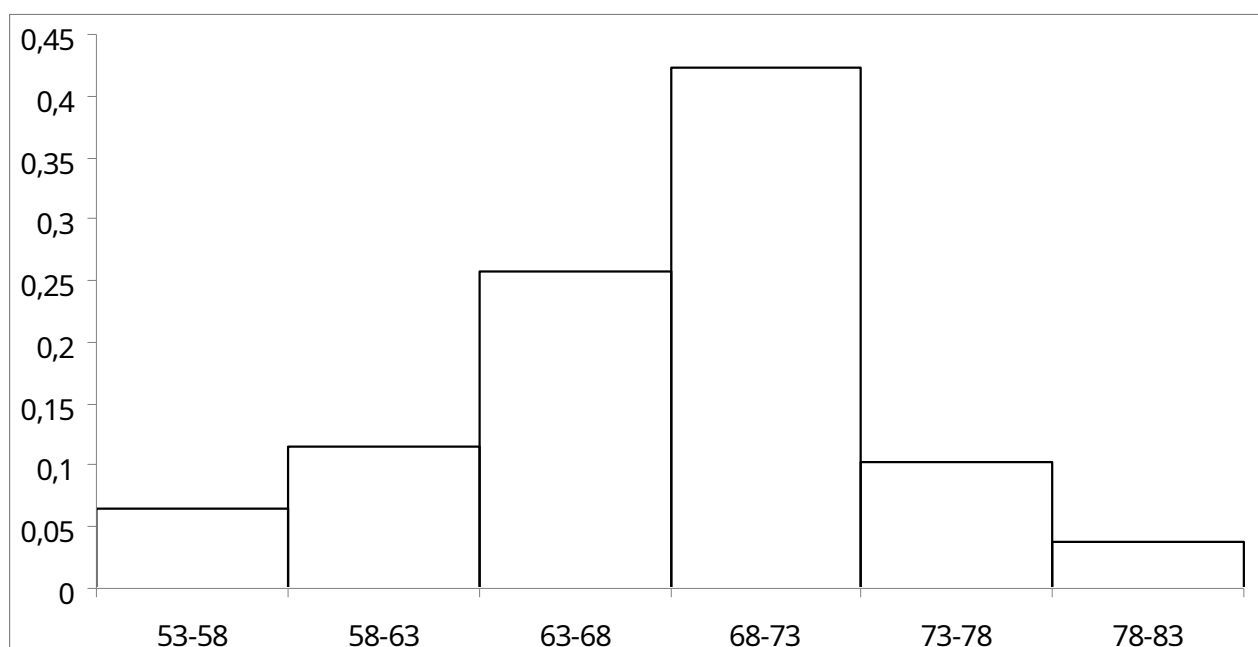
$$n = 78, \quad h_x = 5.$$

Таблица 4

№	$[\beta_{j-1}, \beta_j)$	y_j	n_{y_j}	$\frac{n_{y_j}}{n}$
1	[420, 454)	437	4	0,051
2	[454, 488)	471	9	0,115
3	[488, 522)	505	17	0,218
4	[522, 556)	539	25	0,321
5	[556, 590)	573	18	0,231
6	[590, 624)	607	3	0,038
7	[624, 658)	641	2	0,026

$$n = 78, \quad h_y = 34.$$

4. Построим гистограмму относительных частот и полигон частот для признака X . Гистограмма – ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной $h_x = 5$, а высоты равны плотности относительной частоты $\frac{n_{отн}}{h_x}$. Гистограмма изображена на рис.1. Полигон частот – ломаная линия, соединяющая точки (x_i, n_{x_i}) , изображен на рис 2.

Рис.1. Гистограмма относительных частот признака X

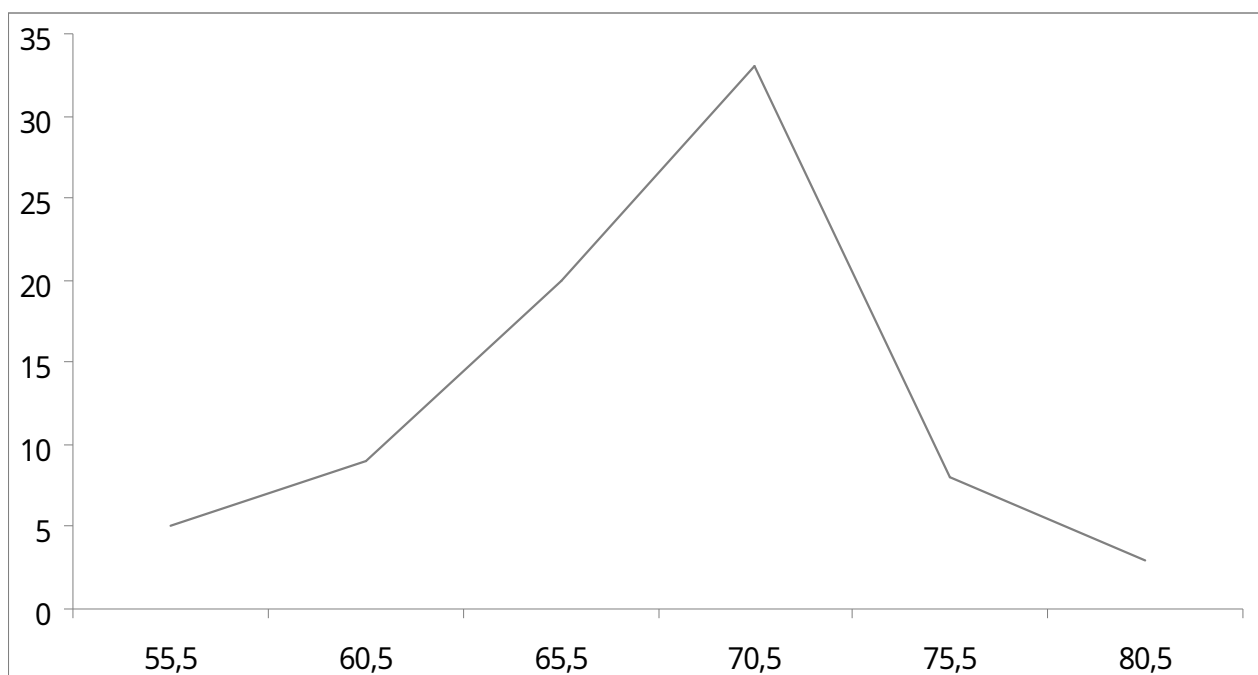


Рис. 2 Полигон частот признака X

Запишем эмпирическую функцию распределения признака X , значения которой уже вычислены в колонке $n_{накоп}$ таблицы 3.

5. Для нахождения числовых характеристик признаков X и Y заполним расчётную таблицу 5.

Таблица 5

№	x_i	$\frac{n_{x_i}}{n}$	$x_i \cdot \frac{n_{x_i}}{n}$	$(x_i - \bar{x}_b) \cdot \frac{n_{x_i}}{n}$
1	55,5	0,064	3,616	10,161
2	60,5	0,115	6,958	6,642
3	65,5	0,257	16,834	1,737
4	70,5	0,423	29,821	2,436
5	75,5	0,103	7,776	5,640
6	80,5	0,0381	3,059	5,843
Σ		1	68,064	32,459

Получим: $\bar{x}_e = 68,1$, $D_{xe} = 32,46$. Аналогично вычисляем числовые характеристики признака Y : $\bar{y}_e = 531,656$, $D_{ye} = 1973,67$. Объем выборки $n = 78$. Подставив это значение в соответствующие формулы для исправленной выборочной дисперсии и исправленного среднеквадратического отклонения, получим:

$$S_x^2 = 32,88, \quad s_x = \sqrt{32,87} = 5,73,$$

$$S_y^2 = 1999,3, \quad s_y = \sqrt{1999,3} = 44,71.$$

Лабораторная работа № 2

Статистическая проверка гипотезы о нормальном распределении признака X генеральной совокупности (X, Y)

Цель работы: проверить по критерию Пирсона согласуется ли гипотеза о нормальном распределении одного из признаков генеральной совокупности при уровне значимости $\alpha = 0,05$. Если нет, то определить уровень значимости, при котором гипотезу о нормальном распределении можно принять. Найти интервальные оценки параметров A и σ с доверительной вероятностью $\gamma = 1 - \alpha$.

Порядок выполнения работы:

Выдвинуть гипотезу H_0 о нормальном распределении признака X .

1. Если частоты первых и последних интервалов малы (меньше пяти [1]), то объединить их с соседними интервалами; l' – новое число интервалов.

2. Вычислить по выборке наблюдаемое значение статистики критерия $\chi^2 = \sum_{i=1}^{l'} \frac{(n_{x_i} - n'_{x_i})^2}{n'_{x_i}}$.

Случайная величина χ^2 имеет χ^2 -распределение Пирсона.

n_{x_i} – наблюдаемая частота, соответствующая i -му интервалу;

n'_{x_i} – теоретическая частота, вычисляемая по формуле $n_{x_i} = n \cdot p_{x_i}$, где

p_{x_i} – теоретическая вероятность попадания случайной величины X в интервал $[\alpha_{i-1}, \alpha_i]$

$p_{x_i} = \Phi\left(\frac{\alpha_i - \bar{x}}{s_x}\right) - \Phi\left(\frac{\alpha_{i-1} - \bar{x}}{s_x}\right)$, где $\Phi(u)$ – функция Лапласа.

3. В таблице χ^2 -распределения Пирсона (приложение 5 [1]) найти критическое значение статистики $\chi^2_{кр}(\alpha; k)$ по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = l' - r - 1$, где l' – число интервалов с учетом их объединения. Для нормального распределения $r = 2$. Сравнить значение $\chi^2_{набл}$, вычисленное по выборке, с табличным критическим значением $\chi^2_{кр}$ и сделать вывод о возможности принятия гипотезы H_0 при уровне значимости $\alpha = 0,05$. Если окажется, что при уровне значимости $\alpha = 0,05$ $\chi^2_{набл} > \chi^2_{кр}$, то по таблице Пирсона найти наименьшее критическое значение статистики $\chi^2_{кр}(\alpha', k)$ и определить уровень значимости α' , при котором можно считать, что данная выборка подчиняется нормальному закону распределения.

4. Найти доверительный интервал для неизвестного математического ожидания A по формуле:

$$\bar{x} - \frac{t_{\gamma} s_x}{\sqrt{n}} \leq a \leq \bar{x} + \frac{t_{\gamma} s_x}{\sqrt{n}}, \quad (1)$$

где t_{γ} – коэффициент Стьюдента, значение которого определяется объемом выборки и надежностью $\gamma = 1 - \alpha_1$ оценки. Значение t_{γ} находим по таблице (приложение 3 [1]).

Доверительный интервал для неизвестного среднеквадратичного отклонения определяется формулой

$$s_x(1 - q) \leq \sigma \leq s_x(1 + q) \quad (2)$$

Значение $q(\gamma, n)$ приведено в таблице (приложение 4 [1]).

5. Записать плотность нормального распределения признака X с параметрами $a = \bar{x}_b$, $\delta = S_x$.

Проверим гипотезу о нормальном распределении X по критерию Пирсона и найдем интервальные оценки A и σ .

- Объединим пятый и шестой интервалы, так как значения частоты в шестом интервале меньше 5 (см. табл. 2).
- Для вычисления по данным выборки значений статистики $\chi^2_{набл}$ составим расчетную таблицу

Таблица 6

№	$\alpha_{i-1} - \alpha_i$	n_{x_i}	$u_i = \frac{\alpha_{i-1} - \bar{x}}{s_x}$	$\Phi(u_i)$	$p'_i = \Delta\Phi$	$n'_{x_i} = n \cdot p'_i$	Целое n'_{x_i}	$\frac{(n_{x_i} - n'_{x_i})^2}{n'_{x_i}}$
1	53 – 58	5	-2,635	-0,4958	0,0347	2,71	3	1,33
2	58 – 63	9	-1,763	-0,4611	0,1478	11,53	11	0,36
3	63 – 68	20	-0,890	-0,3133	0,3063	23,89	23	0,39
4	68 – 73	33	-0,017	-0,0070	0,3107	24,2	25	2,56
5	73 – 83	11	0,855	0,3037	0,1916	14,94	14	0,64
			2,600	0,4953				
Σ		78					76	5,28

При расчетах использовались значения, полученные в лабораторной работе №1: $\bar{x} = 68,1$, $s_x = 5,73$, $n = 78$.

- Наблюдаемое значение статистики $\chi^2_{набл} = 5,28$. По таблице критических точек распределения χ^2 (см. приложение 5 [1]) по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = 2$ находим $\chi^2_{кр}(0,05; 2) = 6,0$. Так как $\chi^2_{набл} < \chi^2_{кр}$, то нет оснований отвергнуть гипотезу H_0 .
- Оценим по формуле (1) неизвестное математическое ожидание при помощи доверительного интервала с надежностью $\gamma = 1 - \alpha' = 1 - 0,025 = 0,975$. При $\gamma = 0,975$, $n = 80$ $t_{\gamma} \approx 2,24$ (см. приложение 3 [1]).

$$68,1 - \frac{2,24 \cdot 5,73}{\sqrt{78}} \leq a \leq 68,1 + \frac{2,24 \cdot 5,73}{\sqrt{78}}$$

$$66,55 \leq a \leq 69,45.$$

По таблице $q(y, n)$ (см. приложение 4 [1]) найдем значение $q = q(0,975; 78) \approx 0,19$.

По формуле (2) находим интервальную оценку параметра σ :

$$5,73 \cdot (1 - 0,19) \leq \sigma \leq 5,73 \cdot (1 + 0,19).$$

$$4,64 \leq \sigma \leq 6,82.$$

5. Для признака X плотность нормального распределения имеем вид

$$f(x) = \frac{1}{5,73 \cdot \sqrt{2\pi}} e^{-\frac{(x-68,1)^2}{2(5,73)^2}}.$$

Строим график функции плотности по гистограмме X .

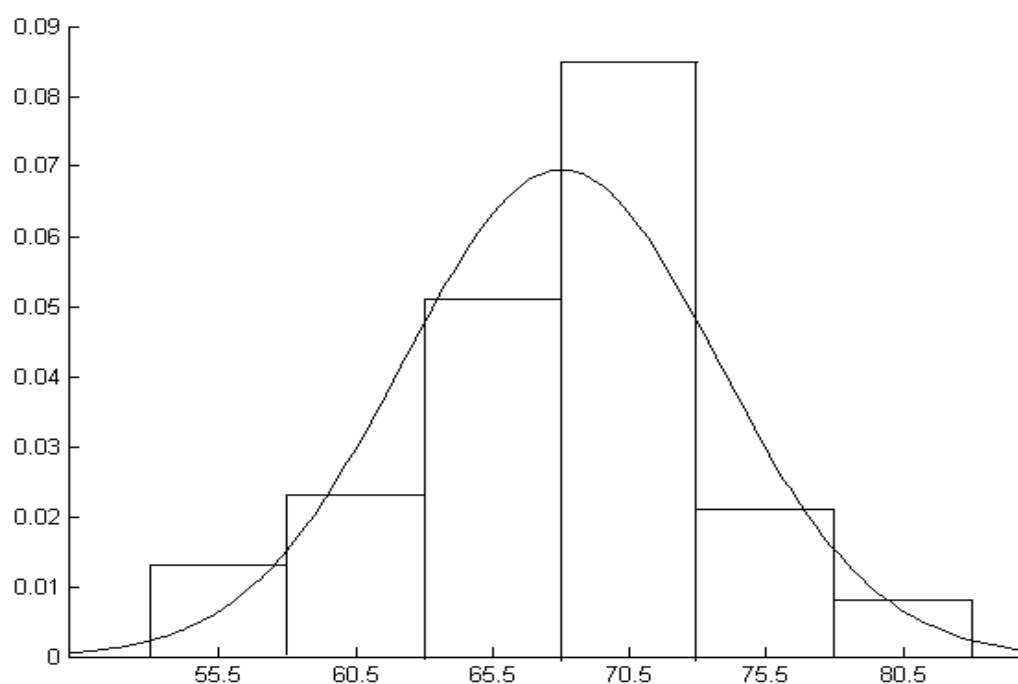


Рис.3. Функция плотности и гистограмма относительных частот

Лабораторная работа №3

Корреляционная зависимость между двумя признаками, построение эмпирической и теоретической линии регрессии

Цель работы: построить эмпирическую линию линейной регрессии, вычислить выборочный коэффициент корреляции, составить теоретическое уравнение линии линейной регрессии и найти доверительный интервал для коэффициента корреляции с надежностью $\gamma = 0,95$.

1. Найти условные средние \bar{Y}_{x_i} и записать их в корреляционную таблицу.
2. Вычислить исправленный выборочный корреляционный момент по формуле

$$S_{xy} = \frac{1}{n-1} \left(\sum_{i=1}^l \sum_{j=1}^m n_{ij} x_i y_j - n \cdot \bar{x} \cdot \bar{y} \right).$$

Найти выборочный коэффициент корреляции

$$r_s = \frac{S_{xy}}{S_x \cdot S_y}.$$

3. Если значение выборочного коэффициента корреляции существенно отличается от единицы, то следует проверить гипотезу H_0 о равенстве нулю генерального коэффициента корреляции Γ при конкурирующей гипотезе $H_1: r \neq 0$. Если нулевая гипотеза будет принята, то выборочный коэффициент незначим, следовательно, X и Y некоррелированы. В качестве критерия проверки

нулевой гипотезы берется значение статистики $T = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}}$. Величина T имеет распределение

Стьюдента с $k = n - 2$ степенями свободы. При $|T| < t_{кр}$ нет оснований опровергнуть гипотезу H_0 , то есть, X и Y некоррелированы. Если $|T| > t_{кр}$, то гипотеза H_0 отвергается. В этом случае следует принять, что коэффициент корреляции существенно отличен от нуля и признаки X и Y коррелированы. Критические точки распределения Стьюдента приведены в таблице (см. приложение 6 [1]).

4. Если выборочный коэффициент корреляции будет признан значимым, то записать уравнение линейной регрессии Y на X .

$$\hat{y}_x - \bar{y} = r_s \frac{S_y}{S_x} (x - \bar{x}).$$

5. Построить в одной системе координат прямую линию регрессии Y на X и эмпирическую линию регрессии Y на X .
6. Найти доверительный интервал для коэффициента корреляции надежностью $\gamma = 0,95$ по формуле:

$$r_s - t_{\gamma} \frac{1 - r_s^2}{\sqrt{n}} < r < r_s + t_{\gamma} \frac{1 - r_s^2}{\sqrt{n}},$$

$t_{\gamma} = t(\gamma, n)$ – коэффициент Стьюдента.

Найдем корреляционную зависимость между X и Y , используя вариационный ряд лабораторной работы №1.

1. Дополним корреляционную таблицу строчкой условных значений признака Y . Для этого вычислим значения признака Y при условии, что X принимает значения x_i :

$$\begin{aligned}\bar{y}_{x=55,5} &= \frac{437 \cdot 4 + 471 \cdot 1}{5} = 443,8; & \bar{y}_{x=70,5} &= \frac{539 \cdot 21 + 573 \cdot 12}{39} = 551,4; \\ \bar{y}_{x=60,5} &= \frac{471 \cdot 8 + 505 \cdot 1}{9} = 474,8; & \bar{y}_{x=75,5} &= \frac{573 \cdot 6 + 607 \cdot 2}{8} = 581,5; \\ \bar{y}_{x=65,5} &= \frac{505 \cdot 16 + 539 \cdot 4}{20} = 511,8; & \bar{y}_{x=80,5} &= \frac{607 \cdot 1 + 641 \cdot 2}{3} = 629,7.\end{aligned}$$

Таблица 7

$Y \backslash X$	55,5	60,5	65,5	70,5	75,5	80,5	n_{y_j}
437	4						4
471	1	8					9
505		1	16				17
539			4	21			25
573				12	6		18
607					2	1	3
641						2	2
n_{x_i}	5	9	20	33	8	3	78
\bar{y}_{x_i}	443,8	474,8	511,8	551,4	581,5	629,7	

2. Для вычисления значения суммы в формуле выборочного корреляционного момента S_{xy} составим расчетную таблицу 8.

Таблица 8

x_i	y_i	n_{ij}	$x_i \cdot y_j \cdot n_{ij}$
55,5	437	4	97014,0
55,5	471	1	26140,5
60,5	471	8	227964,0
60,5	505	1	30552,5
65,5	505	16	529240,0
65,5	539	4	141218,0
70,5	539	21	797989,5
70,5	573	12	484758,0
75,5	573	6	259569,0
75,5	607	2	91657,0
80,5	607	1	48863,5
80,5	641	2	103201,0
Σ			2838167

$$S_{xy} = \frac{1}{78-1} \cdot (2838167 - 78 \cdot 68,1 \cdot 531,656) = 237,19$$

$$r_e = \frac{237,19}{5,73 \cdot 44,71} = 0,93$$

3. Так как полученное значение выборочного коэффициента корреляции $r_e = 0,93$ близко к единице, то X и Y следует признать коррелированными.

4. Запишем уравнение теоретической линии регрессии Y на X :

$$\hat{y}_x - 531,656 = 0,93 \cdot \frac{44,71}{5,73} (x - 68,1) \quad \text{или}$$

$$\hat{y}_x = 7,26x + 38,21.$$

Найдем значения $\hat{y}(x_i)$ и занесем их в таблицу 9

Таблица 9

x_i	55,5	60,5	65,5	70,5	75,5	80,5
\bar{y}_{x_i}	443,8	474,8	511,8	551,4	581,5	629,7
$\hat{y}(x_i)$	441,1	477,4	513,7	550,0	586,3	622,6

Вывод: сравнивая \bar{y}_{x_i} и $\hat{y}(x_i)$, видим, что уравнение регрессии хорошо согласуется с данными выборки.

Построим графики теоретической и эмпирической линий регрессии Y на X рис.3

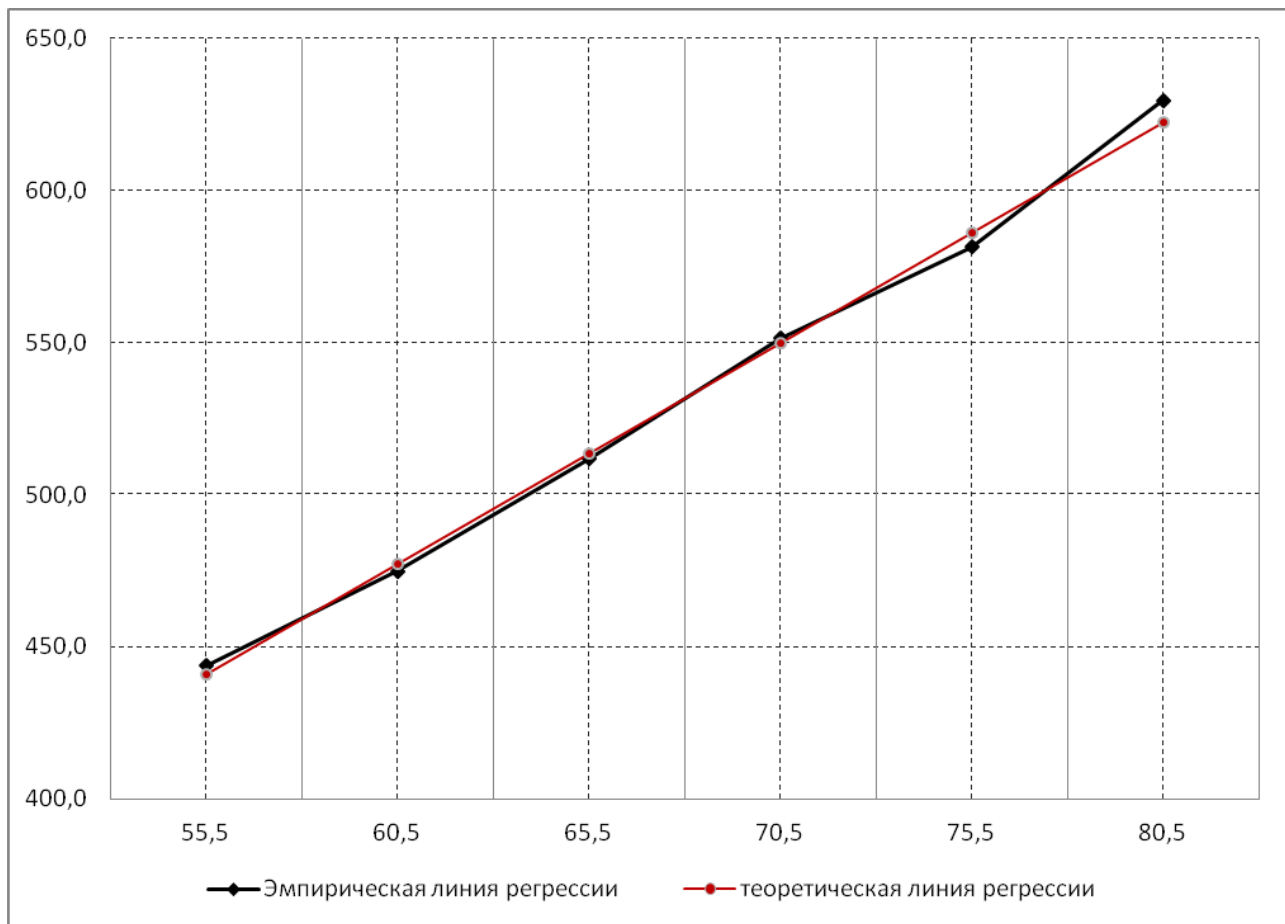


Рис.3 Теоретическая линия регрессии Y на X

5. Определим доверительный интервал для коэффициента корреляции r . Значение $t_{\gamma} = t(\gamma; n)$ найдем по таблице значений коэффициента Стьюдента (см. приложение 3 [1]). Для $\gamma = 0,95$ и $n = 78$ $t_{\gamma} = 1,991$.

$$0,93 - 1,991 \cdot \frac{1 - 0,93^2}{\sqrt{78}} < r < 0,93 + 1,991 \cdot \frac{1 - 0,93^2}{\sqrt{78}},$$

$$0,90 < r < 0,96.$$

Литература

1. Гмурман В.Е., Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высшая школа, 2005.
2. Гмурман В.Е., Теория вероятностей и математическая статистика. – М.: Высшая школа, 2005.
3. Колде Я.К., Практикум по теории вероятностей и математической статистике. – М. Высшая школа, 1991.

Учебное издание

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Составители

Чердынцева Галина Алексеевна,

Кравченко Нелли Михайловна,

Трясцина Татьяна Сергеевна

Корректор -----

Компьютерный набор *Г.А. Чердынцева, Т.С. Трясцина*

Подписано в печать _____.

Формат 60х84 1/16.

Бумага типографская.

Плоская печать.

Усл. печ. л. 1,86.

Уч.-изд. л. 1, 4.

Тираж 50 экз.

Заказ

Редакционно-издательский отдел УрФУ

620002, Екатеринбург, ул. Мира, 19

rio@ustu.ru

Издательско-полиграфический центр УрФУ

620000, г. Екатеринбург, ул. Тургенева, 4