

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

Расчётно-графическая работа по математической статистике

Выполнил: Загвоздин Денис Сергеевич
Группа: РИ-210911
Преподаватель: Поторочина Ксения Сергеевна
Дата: 15.12.2022

Екатеринбург
УрФУ
2022

Список основных обозначений

N — число единиц совокупности

n — число групп (интервалов)

x_{min}, y_{min} — нижняя граница интервала

x_{max}, y_{max} — верхняя граница интервала

h — величина интервала

\bar{x}_e, \bar{y}_e — выборочное среднее

\overline{D}_e — средняя выборочная дисперсия

σ_e — выборочное среднее квадратическое отклонение

S^2 — несмещённая состоятельная оценка генеральной дисперсии

S — исправленное выборочное среднее квадратическое отклонение

Введение

Исследуем корреляцию между оценками студентов в смежных предметах: основы веб-технологий и технологии программирования. Будем брать случайного студента и смотреть его итоговый балл в обеих дисциплинах. X - оценка студента по предмету «Основы веб-технологий». Y - оценка студента по предмету «Технологии программирования».

Вариации дискретные, так как имеют изолированные значения: $X = \{0, 1, 2, \dots, 100\}$; $Y = \{0, 1, 2, \dots, 43\}$.

Материал взят из открытых ведомостей, которые регулярно обновляется, поэтому материал актуален на момент написания работы — 15.12.2022

Цель: исследовать корреляцию между оценками студентов в смежных предметах.

Способ отбора данных: простой случайный отбор.

Ссылки:

[PHP фреймворк, который лёг в основу генерации этого отчёта](#)

[JS библиотека, с помощью которой отображаются LaTeX формулы](#)

[CSS фреймворк, отвечающий за стилизацию страниц](#)

[Руководство по LaTeX, которая позволяет создавать сложные математические формулы](#)

[Ведомость по предмету «Основы веб-технологий»](#)

[Ведомость по предмету «Технологии программирования»](#)

[Исходный код программы](#)

X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
83	13	0	2	5	15	64	7	95	7	30	6
4	8	1	0	0	0	19	16	16	0	30	20
25	26	50	11	8	13	0	3	10	12	0	1
1	2	83	28	59	35	7	3	0	0	24	22
27	36	5	6	0	8	45	5	0	4	100	27
59	27	99	12	43	18	1	4	25	17	46	0
24	6	39	30	0	3	52	16	7	5	27	0
0	1	39	34	34	24	8	11	0	27	0	0
7	21	53	2	17	26	4	10	0	0	33	9
7	10	57	4	94	20	19	35	17	24	100	27
91	24	19	31	0	8	0	2	0	4	95	20
0	0	56	33	1	10	96	23	4	17	45	7
7	15	2	8	14	19	7	8	1	4	28	7
82	14	15	0	8	29	33	7	35	8	1	6
100	34	0	10	49	33	35	33	56	32	79	17
70	4	64	22	40	25	30	34	6	25	27	22
29	32	42	33	26	12	58	39	51	25	56	11
100	38	17	12	4	0	0	1	50	14	12	15
82	31	70	4	73	26	2	3	0	31	45	32
0	1	4	7	0	1	11	7	23	13	56	28
27	20	0	0	100	19	11	12	0	0	4	12
48	14	0	0	100	10	0	8	100	37	94	29
0	0	18	32	0	3	17	26	100	20	99	29
23	8	92	24	100	34	0	1	77	26	0	1
24	26	28	8	29	4	0	0	1	14	0	0

N = 501

X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
0	0	1	6	12	10	2	14	17	4	100	40
39	35	70	29	100	33	0	0	99	17	4	6
0	3	35	22	0	30	0	0	30	8	11	8
68	38	19	17	10	6	0	7	1	21	17	5
0	1	0	1	100	41	100	7	20	26	100	9
1	3	4	8	54	11	85	0	51	39	0	5
100	40	31	6	37	32	0	6	0	2	54	24
17	20	95	42	0	4	100	34	35	3	0	1
0	9	2	10	66	32	0	0	0	0	78	29
0	18	65	29	28	16	0	1	0	2	100	27
48	6	55	36	36	22	2	3	0	11	19	32
8	15	59	11	0	13	75	7	100	7	7	26
34	15	95	33	28	14	100	29	25	17	73	35
0	0	10	0	0	0	7	8	24	34	0	15
54	36	94	39	0	0	73	40	100	39	51	26
0	0	1	12	40	18	5	23	10	21	5	32
23	39	17	42	0	2	0	0	0	4	1	4
5	17	64	39	53	34	0	31	22	22	46	22
0	6	1	8	0	10	5	8	0	7	100	36
100	36	59	21	46	6	0	2	43	33	54	31
25	3	100	25	0	3	100	38	0	0	0	2
0	1	74	29	15	1	15	1	8	4	0	11
0	0	10	3	2	18	8	10	35	34	1	4
75	37	2	27	21	8	17	0	48	12	39	17
35	4	0	4	62	3	44	19	33	29	32	13

N = 501

X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
84	37	1	1	0	1	49	33	21	0	62	4
2	0	28	7	0	0	0	2	57	2	23	15
0	0	19	5	92	35	5	11	23	6	7	16
1	3	79	25	33	11	0	8	1	3	1	7
0	33	0	7	2	2	73	14	0	5	59	5
0	1	52	13	73	27	4	4	2	2	97	36
0	9	24	2	100	26	4	27	94	13	26	28
1	2	0	0	0	5	0	8	39	21	14	36
73	16	10	18	100	25	25	13	100	10	2	25
100	11	2	11	99	10	25	37	57	34	100	33
0	0	100	39	0	7	100	4	95	1	1	15
0	0	42	7	60	30	1	35	0	0	13	3
0	12	1	0	2	36	7	6	0	6	2	11
1	3	28	30	0	11	95	5	0	13	2	14
100	9	6	0	7	7	2	31	61	41	8	7
54	19	84	18	0	6	1	8	0	0	48	15
65	5	2	1	18	27	100	32	100	34	1	11
0	6	98	34	2	0	100	13	100	37	8	12
8	7	0	5	58	13	4	2	0	6	45	34
87	29	0	3	0	12	8	30	0	0	0	0
46	20	0	0	4	6	0	0	0	5	18	24
10	14	2	28	29	7	100	32	0	10	0	3
0	4	8	1	24	9	40	12	100	13	100	44
0	7	67	36	94	20	24	4	18	10	23	4
30	3	27	33	100	38	0	6	1	7	100	40

N = 501

X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
24	13	1	2	5	10	100	21	12	3	38	22
0	3	16	8	100	28	24	27	7	4	100	21
0	0	2	32	0	4	22	11	1	8	100	36
4	5	2	17	46	28	2	21	95	2	13	25
2	18	64	27	20	15	0	2	1	5	100	34
24	15	25	42	1	0	0	11	5	6	0	5
15	24	45	0	63	38	4	9	18	3	2	11
0	4	100	33	100	4	40	11	4	4	0	1
18	8	23	37	0	4						
N = 501											

Первичная обработка результатов эксперимента

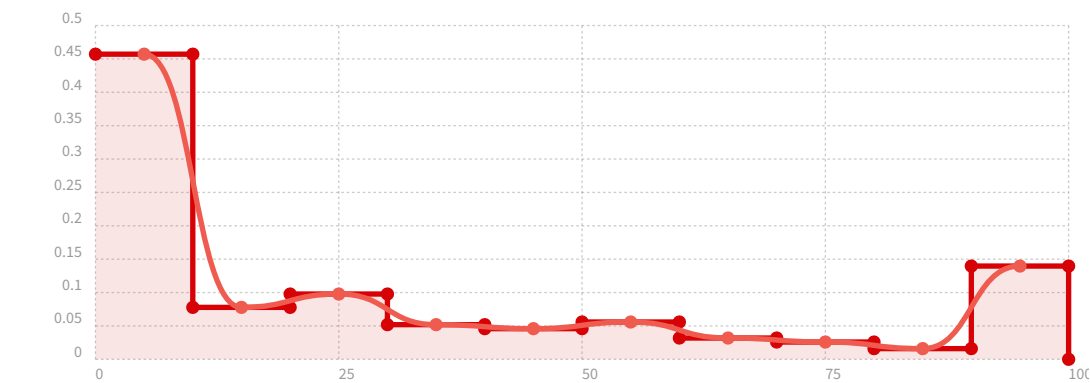
Обработка одномерной СВ X .

$$X = \{x_i\}$$
$$x_{min} = 0 \text{ (1)}; m = 119$$
$$x_{max} = 100 \text{ (99)}; m = 48$$
$$x_{max} - x_{min} = 100 - 0 = 100 - \text{размах выборки}$$
$$n = 1 + 3.22 \lg N = 1 + 3.22 \frac{\lg N}{\lg 10} = 1 + 3.22 \frac{\lg 501}{\lg 10} = 9.6934774772925 \approx 10 - \text{число интервалов}$$
$$h = \frac{x_{max} - x_{min}}{n} = \frac{100 - 0}{10} = 10 - \text{число групп (интервалов)}$$

Построим статистический ряд.

$(x_{i-1}; x_i]$	n_i	$\omega_i = \frac{n_i}{N}$	$x_{i \text{ сеп}}$
[0; 10]	229	0.4571	5
(10; 20]	39	0.0778	15
(20; 30]	49	0.0978	25
(30; 40]	26	0.0519	35
(40; 50]	23	0.0459	45
(50; 60]	28	0.0559	55
(60; 70]	16	0.0319	65
(70; 80]	13	0.0259	75
(80; 90]	8	0.016	85
(90; 100]	70	0.1397	95

Гистограмма и полигон частот



Найдём числовые характеристики

$$\bar{x}_e = \sum_{i=1}^{10} x_i \omega_i = 5 \cdot 0.4571 + 15 \cdot 0.0778 + 25 \cdot 0.0978 + 35 \cdot 0.0519 + 45 \cdot 0.0459 + 55 \cdot 0.0559 + 65 \cdot 0.0319 + 75 \cdot 0.0259 + 85 \cdot 0.016 + 95$$
$$\overline{D}_e = \sum_{i=1}^{10} x_i^2 \omega_i - (\bar{x}_e)^2 = 5^2 \cdot 0.4571 + 15^2 \cdot 0.0778 + 25^2 \cdot 0.0978 + 35^2 \cdot 0.0519 + 45^2 \cdot 0.0459 + 55^2 \cdot 0.0559 + 65^2 \cdot 0.0319 + 75^2 \cdot 0.0259$$
$$\sigma_e = \sqrt{\overline{D}_e} = \sqrt{1080.193} = 32.8663$$

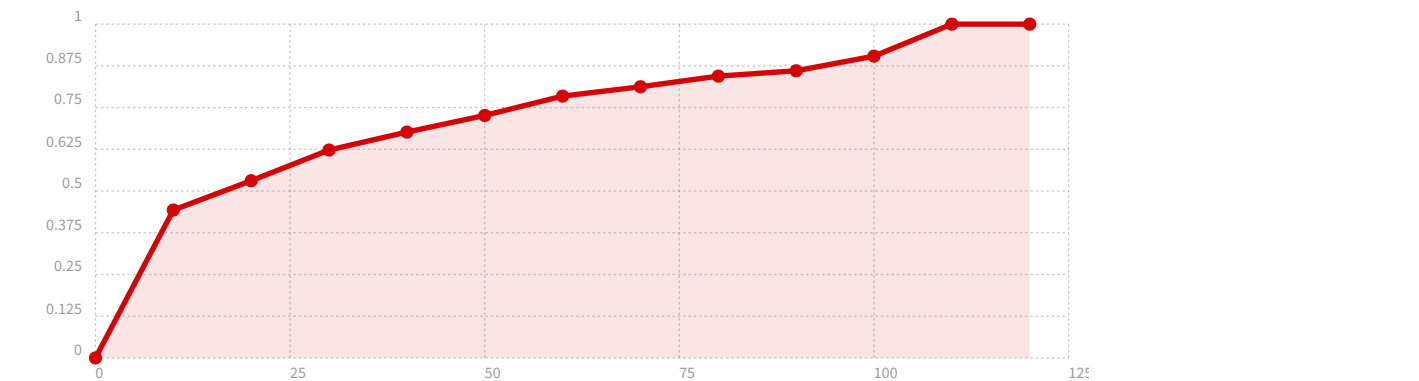
$$S^2 = \frac{N}{N-1} \overline{D}_e = \frac{501}{501-1} 1080.193 = 1082.3534$$
$$S = \sqrt{S^2} = \sqrt{1082.3534} = 32.8991$$
$$A_S = \frac{\mu^3}{S^3} = \frac{\sum_{i=1}^{10} (x_i - \overline{x}_e)^3 \omega_i}{S^3} = \text{много} - \text{коэффициент асимметрии}$$
$$E_S = \frac{\mu^4}{S^4} - 3 = \frac{\sum_{i=1}^{10} (x_i - \overline{x}_e)^4 \omega_i}{S^4} - 3 = \text{много} - \text{коэффициент эксцесса}$$

Коэффициент асимметрии характеризует меру скошенности графика/вправо, а эксцесс — меру его высоты. То, что эти коэффициенты принимают большие значения, ожидаемо. Из полигона частот мы уже могли сделать вывод, что распределение не является нормальным. Большие значения коэффициентов помогают лишний раз убедиться, что распределение отличное от нормального. На перёд — для СВ Y коэффициенты тоже принимают большие значения, что позволяет сделать аналогичный вывод.

Найдём эмпирическую функцию распределения и построим ее график.

$(x_{i-1}; x_i]$	x	$F^*(x) = W(X < x)$
$(-\infty; 0]$	0	0
$(0; 10]$	10	0.4431
$(10; 20]$	20	0.5309
$(20; 30]$	30	0.6228
$(30; 40]$	40	0.6766
$(40; 50]$	50	0.7265
$(50; 60]$	60	0.7844
$(60; 70]$	70	0.8124
$(70; 80]$	80	0.8443
$(80; 90]$	90	0.8603
$(90; 100]$	100	0.9042
$(100; +\infty)$	$+\infty$	1

Кумулята:



Проверка критерия Пирсона

Установим уровень значимости α , при котором распределение для выборки согласуется с **показательным законом** распределения по критерию Пирсона (χ^2). Возьмём готовые формулы, которые мы получили на практиках. Функция плотности для показательного распределения:

$$f(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

$$p_i(x \in (x_{i-1}; x_i)) = e^{-\frac{x_{i-1}}{x_0}} - e^{-\frac{x_i}{x_0}}$$

Выдвинем гипотезу: $H_0 : X \sim P(\lambda)$

Выдвинем противоположную гипотезу: $H_1 : X$ — имеет отличное от показательного распределение

\mathcal{N}_i^2	$x_{i-1} - x_i$	n_i	p_i	$n'_i = p_i N$	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
1	0—10	229	0.272	136	93	8649	63.5956
2	10—20	39	0.198	99	−60	3600	36.3636
3	20—30	49	0.1442	72	−23	529	7.3472
4	30—40	26	0.1049	53	−27	729	13.7547
5	40—50	23	0.0764	38	−15	225	5.9211
6	50—60	28	0.0556	28	0	0	0
7	60—70	16	0.0405	20	−4	16	0.8
8	70—80	13	0.0295	15	−2	4	0.2667
9	80—90	8	0.0215	11	−3	9	0.8182
10	90—100	70	0.0156	8	62	3844	480.5
		501					$\chi^2_{набл} = 609.3671$

Число степеней свободы: $r = k - m - 1 = 10 - 1 - 1 = 8$

- $\chi^2_{крит} = 13.362$; при $\alpha = 0.100$
- $\chi^2_{крит} = 15.507$; при $\alpha = 0.050$
- $\chi^2_{крит} = 17.535$; при $\alpha = 0.025$
- $\chi^2_{крит} = 18.168$; при $\alpha = 0.020$
- $\chi^2_{крит} = 20.090$; при $\alpha = 0.010$
- $\chi^2_{крит} = 21.955$; при $\alpha = 0.005$
- $\chi^2_{крит} = 26.124$; при $\alpha = 0.001$

Вывод: ни при каком уровне значимости α распределение X не подчиняется показательному распределению. Почему так происходит? Из графика полигона частот мы можем предположить, что наиболее подходящее распределение — показательное распределение. Так как график плавно убывает, что очень напоминает убывающую показательную функцию. Но гипотеза H_0 не подтверждается. Из таблицы мы видим, что у нас **очень** тяжёлые концы, которы и составляют большую часть $\chi^2_{набл}$. А интервалы под номерами 6, 7, 8 и 9 практически не накапливают ошибку — идут почти вровень с показательными распределением.

Обработка одномерной СВ Y .

- $Y = \{y_i\}$
- $y_{min} = 0$ (1); $m = 51$
- $y_{max} = 44$ (42); $m = 1$

$$y_{\max} - y_{\min} = 44 - 0 = 44 - \text{размах выборки}$$

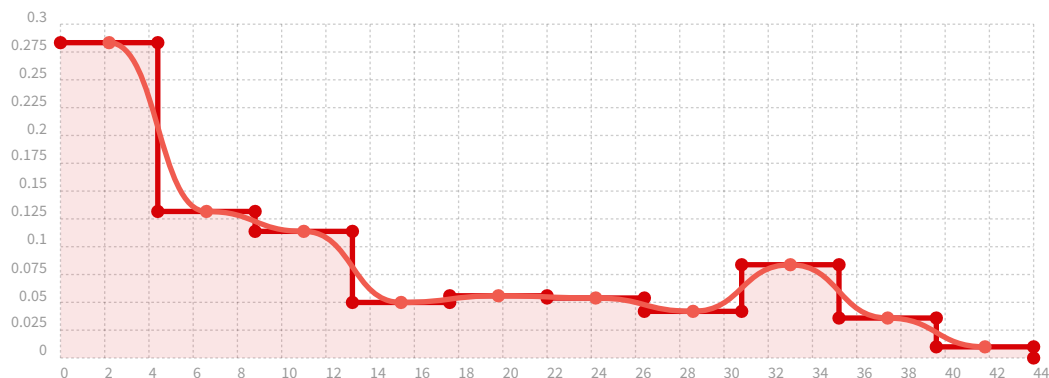
$$n = 1 + 3.22 \lg N = 1 + 3.22 \frac{\lg N}{\lg 10} = 1 + 3.22 \frac{\lg 501}{\lg 10} = 9.6934774772925 \approx 10 - \text{число интервалов}$$

$$h = \frac{y_{\max} - y_{\min}}{n} = \frac{44 - 0}{10} = 4.4 - \text{число групп (интервалов)}$$

Построим статистический ряд.

$(y_{i-1}; y_i]$	n_i	$\omega_i = \frac{n_i}{N}$	$y_{i \text{ сеп}}$
$[0; 4.4]$	142	0.2834	2.2
$(4.4; 8.8]$	66	0.1317	6.6
$(8.8; 13.2]$	57	0.1138	11
$(13.2; 17.6]$	25	0.0499	15.4
$(17.6; 22]$	28	0.0559	19.8
$(22; 26.4]$	27	0.0539	24.2
$(26.4; 30.8]$	21	0.0419	28.6
$(30.8; 35.2]$	42	0.0838	33
$(35.2; 39.6]$	18	0.0359	37.4
$(39.6; 44]$	5	0.01	41.8

Гистограмма и полигон частот



Найдём числовые характеристики

$$\bar{y}_e = \sum_{i=1}^{10} y_i \omega_i = 2.2 \cdot 0.2834 + 6.6 \cdot 0.1317 + 11 \cdot 0.1138 + 15.4 \cdot 0.0499 + 19.8 \cdot 0.0559 + 24.2 \cdot 0.0539 + 28.6 \cdot 0.0419 + 33 \cdot 0.0838 + 37.4 \cdot 0.0359 + 41.8 \cdot 0.01 = 11.9884$$

$$\overline{D}_e = \sum_{i=1}^{10} y_i^2 \omega_i - (\bar{y}_e)^2 = 2.2^2 \cdot 0.2834 + 6.6^2 \cdot 0.1317 + 11^2 \cdot 0.1138 + 15.4^2 \cdot 0.0499 + 19.8^2 \cdot 0.0559 + 24.2^2 \cdot 0.0539 + 28.6^2 \cdot 0.0419 + 33^2 \cdot 0.0838 + 37.4^2 \cdot 0.0359 + 41.8^2 \cdot 0.01 - 11.9884^2 = 143.7224$$

$$\sigma_e = \sqrt{\overline{D}_e} = \sqrt{143.7224} = 11.9884$$

$$S^2 = \frac{N}{N-1} \overline{D}_e = \frac{501}{501-1} 143.7224 = 144.0098$$

$$S = \sqrt{S^2} = \sqrt{144.0098} = 12.0004$$

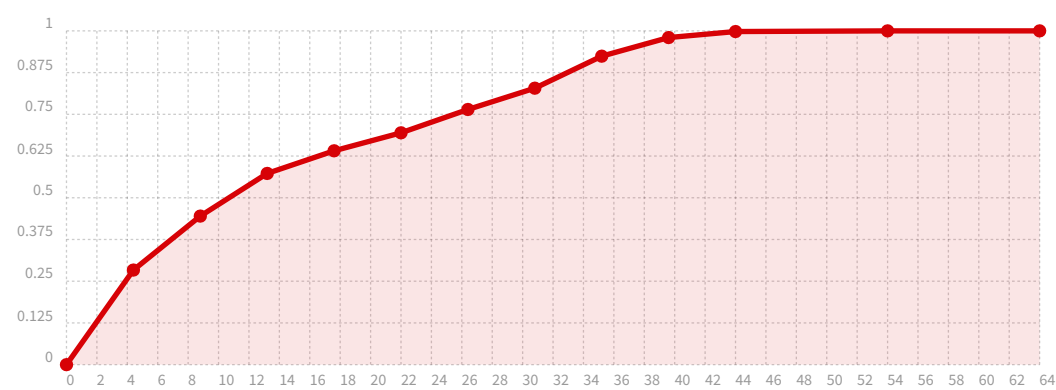
$$A_S = \frac{\mu_3^3}{S^3} = \frac{\sum_{i=1}^{10} (y_i - \bar{y}_e)^3 \omega_i}{S^3} = \text{много} - \text{коэффициент асимметрии}$$

$$E_S = \frac{\mu_4^4}{S^4} - 3 = \frac{\sum_{i=1}^{10} (y_i - \bar{y}_e)^4 \omega_i}{S^4} - 3 = \text{много} - \text{коэффициент эксцесса}$$

Найдём эмпирическую функцию распределения и построим ее график.

$(y_{i-1}; y_i]$	y	$F^*(y) = W(Y < y)$
$(-\infty; 0]$	0	0
$(0; 4.4]$	4.4	0.2834
$(4.4; 8.8]$	8.8	0.4451
$(8.8; 13.2]$	13.2	0.5729
$(13.2; 17.6]$	17.6	0.6407
$(17.6; 22]$	22	0.6946
$(22; 26.4]$	26.4	0.7645
$(26.4; 30.8]$	30.8	0.8283
$(30.8; 35.2]$	35.2	0.9242
$(35.2; 39.6]$	39.6	0.98
$(39.6; 44]$	44	0.998
$(44; +\infty)$	$+\infty$	1

Кумулята:



Двумерная выборка:

X\Y	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	44	
0	35	15	9	8	9	6	7	5	5	2	3	4	2	2		1			1										1			1	2		1										
1	3	1	3	4	4	1	2	2	3		1	1	1		1	1						1														1									
2	2	1	2	2					1		1	3			2			1	2			1				1		1	1			1	1				1								
4	1		1		2	1	2	1	2	1	1		1					1											1																
5						2		1		1	1					1		1						1											1										
6	1																										1																		
7				1	1	1	1	1	2		1					1	1					1						1																	
8		1			1			2			1	1	1	1		1															1	1													
10	1			1			1						1		1				1			1																							
11								1	1				1																																
12				1							1					1																													
13				1																							1																		
14																				1																						1			
15	1	2																								1																			

13/14

14/14