МИНИСТЕРСТВО НАУКИ И ОБРАЗОВАНИЯ РФ

Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский университет
ИТМО»

ФАКУЛЬТЕТ ПРОГРАММНОЙ ИНЖЕНЕРИИ И КОМПЬЮТЕРНОЙ ТЕХНИКИ

РАСЧЕТНО-ГРАФИЧЕСКАЯ РАБОТА ПО ТЕМЕ «ИНТЕГРАЛ ФУНКЦИИ ОДНОЙ ПЕРЕМЕННОЙ»

по дисциплине

Математическая статистика

Вариант № 3

Выполнили:
Карташев Владимир, Р3215
Карабанов Андрей, Р3215
Преподаватель:
Кудрявцева Ирина Владимировна

г. Санкт-Петербург

2024 год

1. Постановка задачи.

Дан статический материал: 120 измерений диаметра цапф, которые дали следующие отклонения от номинального размера (в миллиметрах):

48	44	39	43	41	46	45	43	44	42	30	48
49	39	37	45	40	36	42	48	33	51	42	38
41	34	42	42	40	48	50	39	47	43	34	40
48	30	42	41	32	30	39	51	50	31	32	39
38	36	44	41	37	33	37	37	25	42	40	43
33	35	49	43	34	40	35	46	35	38	43	44
34	36	43	45	42	41	40	34	44	37	42	43
52	45	39	35	39	43	46	37	40	36	45	51
32	36	41	31	32	43	34	41	44	34	40	34
33	49	43	34	49	47	36	30	48	45	32	46

Используя статистический материал (результаты измерений), необходимо:
1) определить размах варьирования R;
2) составить статистический ряд распределения частот СВ X;
3) составить интервальные статистические ряды частот и относительных частот;
4) построить полигон и гистограмму относительных частот;
5) найти эмпирическую функцию распределения и построить её график;
6) вычислить выборочные значения числовых характеристик СВ X: математического ожидания M(X) , дисперсии Д(X) , среднего квадратического

отклонения σ (X);
7) по виду полигона и гистограммы относительных частот сделать выбор закона распределения СВ X;
8) найти точечные оценки параметров предполагаемого распределения, записать функцию распределения и плотность вероятности СВ X;
9) проверить согласие эмпирической функции распределения F*(x) с теоретической F(x) при помощи критерия согласия 2—Пирсона или λ — Колмогорова.
В случае нормального распределения СВ X по заданному уровню значимости a:

1) Найти интервальные оценки параметров распределения;
2) Проверить нулевую гипотезу о математическом ожидании при альтернативной гипотезе

2. Выполнение задачи

2.1 Определение размаха варьирования R.

X_{min} = 25
X_{max} = 52
R=X_{max} - X_{min} = 52 - 25 = 27 - размах варьирования (разность между крайними значениями вариант вариационного ряда), т.е. возможные значения СВ X (диаметр цапф) принадлежат отрезку [25;52]

2.2. Составление статистического ряда распределения частот СВ X.

Составим статистический ряд распределения частот (таблицу 2).

Диаметр цапф xi, мм	Частота mi
25	1
30	4
31	2
32	5
33	4
34	9
35	4
36	6
37	6
38	3
39	7
40	8
41	7
42	9
43	11

44	6
45	6
46	4
47	2
48	6
49	4
50	2
51	3
52	1
	120

Таблица 2 - Вариационный ряд (столбец 1). Статистический ряд распределения частот (столбцы 1 и 3)

Контроль: $\sum_i x_i$ = 120

2.3 Составление интервальных статистических рядов частот и относительных частот.

$h = \frac{R}{1 + 3{,}2\lg n}$, ,где h - длина интервала, n = 120 объем выборки.

Число интервалов k равно округленному до целого R/h.

h	3,527853059
k	7,653379987
h*	4,5

Составим интервальные статистические ряды распределения частот и относительных частот (таблица 3, стб. 2;4 и 2;5). В столбцах 6-10 подсчитаны числа, нужные при выполнении следующих пунктов задачи.

2.4 Построение полигона и гистограммы относительных частот.

№	Диапазон		Частота в интервале	Относ. частота wi	Накопленная частота nx	Относительная накопленная частота	Плотность частоты	Плотность относительной частоты	Середина интервала
	от	до							
1	25	29,5	1	0,008	1	0,008	0,222	0,002	27,25
2	29,5	34	15	0,125	16	0,133	3,333	0,028	31,75
3	34	38,5	28	0,233	44	0,367	6,222	0,052	36,25
4	38,5	43	31	0,258	75	0,625	6,889	0,057	40,75

5	43	47,5	29	0,242	104	0,867	6,444	0,054	45,25
6	47,5	52	16	0,133	120	1,000	3,556	0,030	49,75
SUM			120	1,000	240	3,000	26,667	0,222	231

Таблица 3 - Интервальные статистические ряды

Контроль: $\sum_{i}^{6} n_i = 120$, $\sum_{i}^{6} w_i = 1$

Построим полигон относительных частот (рисунок 1) по данным столбцов 5 и 10 (см. таблицу 3).

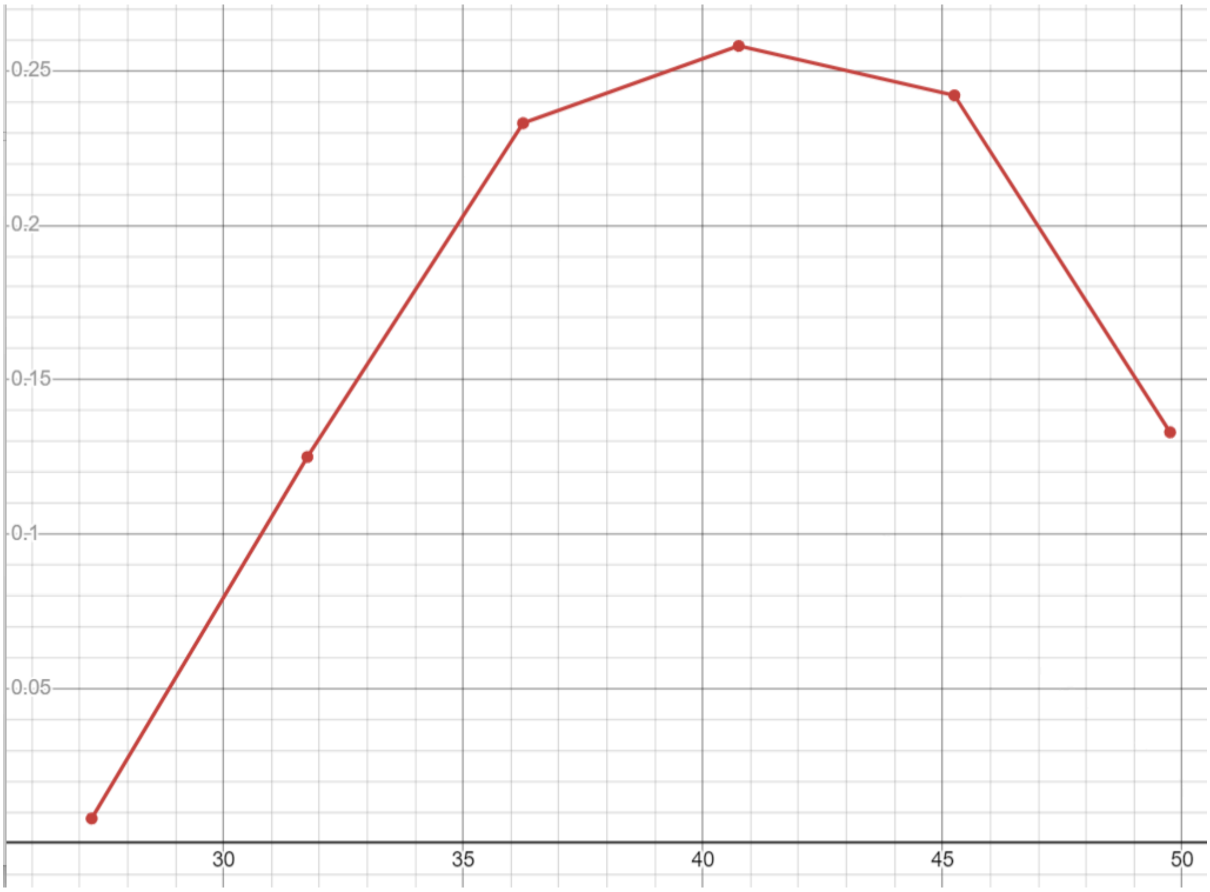


Рисунок 1

Площадь гистограммы относительных частот равна $\sum_{i}^{6} w_i = 1$.

По гистограмме и полигону относительных частот можно судить о форме эмпирической кривой распределения — графике функции $f^*(x)$ (эмпирической плотности вероятности).

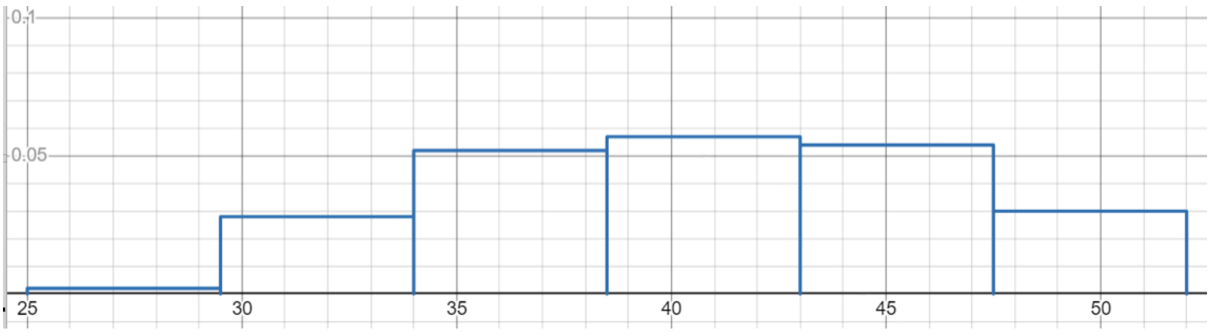


Рисунок 2

2.5 Нахождение эмпирической функции распределения и построение её графика

По данным таблицы 3 (стб: 2, 7) найдем эмпирическую функцию распределения

x	<=25	29,5	34	38,5	43	47,5	>52
F*(x)	0	0,008	0,133	0,367	0,625	0,867	1,000

Таблица 4 - Эмпирическая функция распределения

Построим график $F^{*}(x)$: сначала на интервалах $(-\infty;13,32)$ и $(13,62;\infty)$, а затем в указанных в таблице 4 точках. Учитывая непрерывность функции $F^{*}(x)$, полученные точки соединим (рисунок 3).

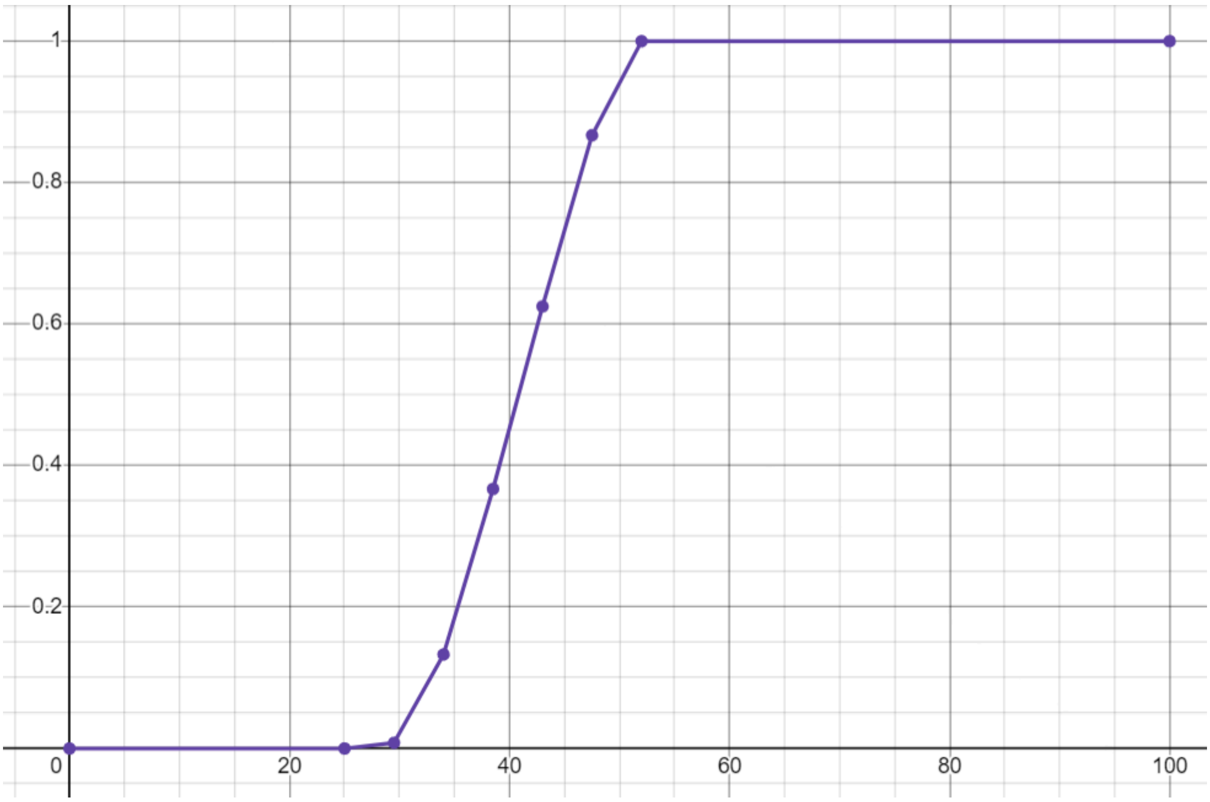


Рисунок 3

2.6 Вычисление выборочных значений числовых характеристик СВ X.

3.6 Находим $M_в(X) = \overline{X}$ по формуле

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{6} y_i \cdot n_i ,$$

где y_i — середина i-го интервала;

n_i — его частота;

n — объём выборки (таблица 3, стб. 4, 10).

Вычислим $Д_в(X)$ по формуле

$$Д_в(X) = \frac{1}{n}\sum_{i=1}^{6}\left(y_i - \overline{X}\right)^2 \cdot n_i .$$

i	yi	ni	yi * ni	(yi - M(X))^2 * ni
1	27,25	1	27,25	182,25
2	31,75	15	476,25	1215
3	36,25	28	1015	567
4	40,75	31	1263,25	0
5	45,25	29	1312,25	587,25
6	49,75	16	796	1296
SUM	231	120	4890	3847,5

Вычисленные значения	
M(X)	40,75
D(X)	32,06
σ(X)	5,66

2.7 Выбор закона распределения СВ X по виду полигона и гистограммы относительных частот.

Полигон и гистограмма относительных частот (см. рисунки 1 и 2) напоминают нормальную кривую (рисунок 4, кривая Гаусса). Поэтому *предположим*, что распределение СВ X(диаметра головки заклепки) является *нормальным*.

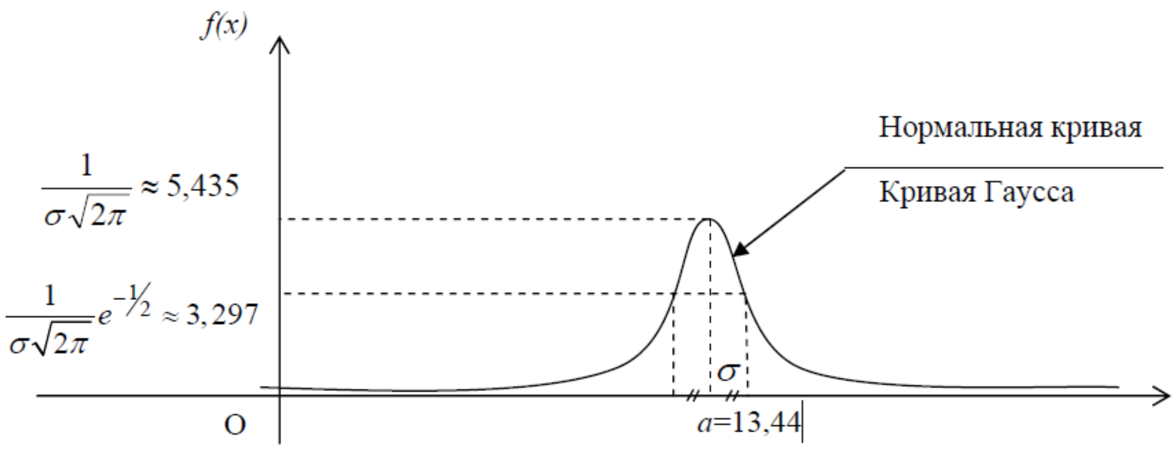


Рисунок 4. Кривая Гаусса

2.8 Нахождение точечной оценки параметров предполагаемого распределения, запись функции распределения и плотности вероятности СВ X:

Плотность вероятности и функция распределения СВ X, распределённой по нормальному закону, имеют вид:

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

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(x-a)^2}{2\sigma^2}}dt.$$

Найдем точечные оценки параметров a = M(X) и σ =σ (X) нормального распределения:

Вспомогательные величины	
~a	40,24166667
σ	5,66

Следовательно, плотность вероятности предполагаемого распределения N(a,σ) имеет вид:

Вычисленные значения	
f(x)	формула
F(x)	формула

её график изображен на рисунке 4.

2.9 Проверка согласия эмпирической функции распределения с теоретический при помощи критерия Колмогорова.

Проверим гипотезу о нормальном распределении с помощью критерия согласия – Колмогорова. Все вспомогательные расчеты, необходимые для нахождения выборочной характеристики, сведем в таблицу 5. Таблица 5 — Нахождение выборочного значения.

№	Диапазон		Частота ni	nx	pi	F*(x) = nx/n	F(x) = P(X < x)	|F*(x) - F(x)|
	от	до						
1	-Inf	29,5	1	1	0,0289	0,008	0,0289	0,0206
2	29,5	34	15	16	0,1063	0,133	0,1352	0,0018
3	34	38,5	28	44	0,4856	0,367	0,6208	0,2541
4	38,5	43	31	75	0,0661	0,625	0,6869	0,0619
5	43	47,5	29	104	0,2131	0,867	0,9001	0,0334

6	47,5	Inf	16	120	0,0999	1,000	1,0000	0,0000
SUM			120	-	1,0000	-	-	-

Таблица - Нахождение выборочного значения λ

Вычисленные значения	
Д	0,2541
λ набл	2,78390493
P	0,95
λ 0,05	1,358

2.10 Нахождение интервальных оценок параметров распределения.

Доверительные интервал, накрывающий математическое ожидание СВ X с надежностью P=1 - a имеет вид:

$$\tilde{a} - t\frac{\sigma_в}{\sqrt{n}} < a < \tilde{a} + t\frac{\sigma_в}{\sqrt{n}}.$$

По таблице квантилей распределения Стьюдента [1,приложения] по заданному уровню значимости a=0.05 и числу степеней свободы v=n-1=89 найдём квантиль

t=1,9802

Вычислим точность оценки:

$$\delta = t \cdot \frac{\sigma_в}{\sqrt{n}}$$

σ=1,0236

Искомый доверительный интервал для M(X) = a

39,73<a<=41,77

Полуинтервал (39,73; 41,77] накрывает неизвестное M(x) с вероятностью P=0,95.

Доверительный интервал, накрывающий среднее квадратическое отклонение СВ X с надежностью P=1-a:

$$\sqrt{\frac{n-1}{\chi^2_{\frac{\alpha}{2},\ v}}} \cdot \sigma_в < \sigma < \sqrt{\frac{n-1}{\chi^2_{1-\frac{\alpha}{2},\ v}}} \cdot \sigma_в$$

или короче

$$j_1 \cdot \sigma_в < \sigma < j_2 \cdot \sigma_в \ ,$$

где $j_1 = \sqrt{\frac{n-1}{\chi^2_{\frac{\alpha}{2},v}}}, \qquad j_2 = \sqrt{\frac{n-1}{\chi^2_{1-\frac{\alpha}{2},v}}}.$

По таблице распределения X^2 по заданной доверительной вероятности p=0,95 и числу степеней свободы v=89 найдем числа j_1=0,8875 и j_2=1,1454

Искомый доверительный интервал для параметра σ:

(5,0253<σ<6,4859]

Полуинтервал (5,0253;6,4859] накрывает неизвестное σ(x) с вероятностью p=0,95.

2.11 Проверка гипотезы H_0: $a = a_0$ и альтернативной гипотезы H_0: $a < a_0$

Правило 3. Пусть $H_0 : a = a_0$ и альтернативная гипотеза $H_a : a < a_0$. По правилу 2 находят вспомогательное критическое значение $U_{кр}$, полагают $U^*_{кр.} = -U_{кр.}$ (граница левосторонней критической области). Если $U_{набл.} > -U_{кр.}$, H_0 принимают; если $U_{набл.} < -U_{кр.}$, H_0 отвергают

24

(наблюдаемое значение U-критерия попадает в критическую область) и принимают альтернативную гипотезу H_a.

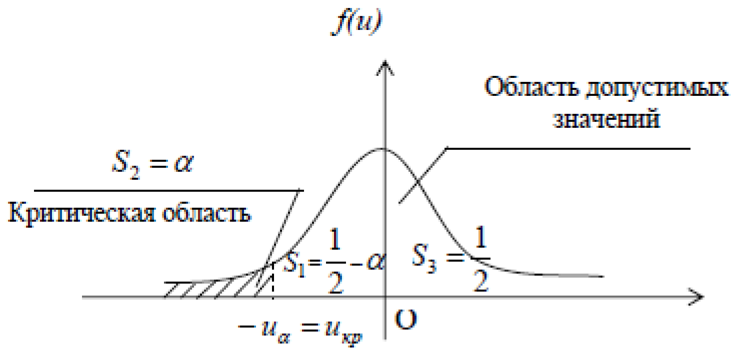


Рисунок 7

Уровень значимости a=0,05
Из таблицы имеем $a_0=a_2=39{,}73$. Надо проверить нулевую гипотезу, против альтернативной.
$\Phi(U_{кр})=0{,}45$, $U_{кр}=1{,}645$, $U_{набл} = 1{,}715$
$U_{кр} > U_{набл}$ => Гипотеза H_0 принимается

2.12 Проверка гипотезы H_0: $\sigma^2 = \sigma^2_0$ и альтернативной гипотезы H_0: $\sigma^2 \mathrel{!=} \sigma^2_0$

Правило 5. Пусть $H_0: \sigma^2 = \sigma_0^2$ и альтернативная гипотеза $H_a: \sigma^2 \neq \sigma_0^2$. Находят критические левую и правую точки $\chi^2_{лев.кр.} = \chi^2_{1-\frac{\alpha}{2},\nu}$ и $\chi^2_{прав.кр.} = \chi^2_{\frac{\alpha}{2},\nu}$ по таблице распределения χ^2—Пирсона.

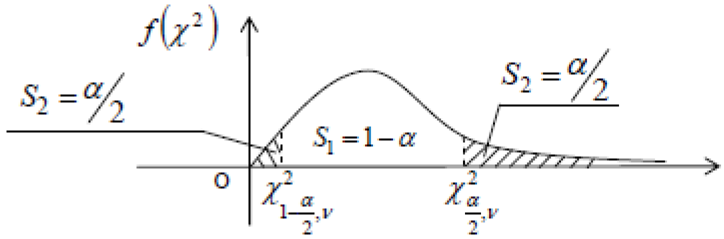


Рисунок 9

Если $\chi^2_{лев.кр.} < \chi^2_{набл.} < \chi^2_{прав.кр.}$ (наблюдаемое значение критерия попало в область допустимых значений), то нулевую гипотезу H_0 принимают. Если $\chi^2_{набл.} < \chi^2_{лев.кр.}$ или $\chi^2_{набл.} > \chi^2_{прав.кр.}$, то отклоняют H_0 в пользу альтернативной гипотезы H_a.

$\sigma_0=\sigma_1=6{,}4859$

$X^2_{лев.кр.}=90{,}67$, $X^2_{прав.\ кр..}=90{,}67$, $X^2_{набл.}=90{,}67$

$X^2_{лев.кр.} <$ $X^2_{прав.\ кр..} <$, $X^2_{набл.}=90{,}67$ => Гипотеза H0 принимается.

3. МНК

3.1 Задание 1

В таблице приведены данные о расходе топлива (y, л на 100 км) автомобиля с двигателем объемом 1,5 литра с автоматической трансмиссией в зависимости от скорости движения (x, км/ч).

x_i	10	20	40	60	90	110	130	140	150	160
y_i	3,8	4	4,2	4,8	5,5	6	7	8,1	10	12

В предположении, что между x и y существует линейная зависимость, определить параметры линейной регрессии

23

$y = kx + b$ методом наименьших квадратов. Спрогнозировать расход топлива при скорости 170 км/ч.

y = kx + b

170 - ?

1. Введем в таблицу согласно варианта эмпирические данные. Произведем необходимые вычисления сумм.

№	xi	yi	x^2	x*y	y рас.i	δ	δ^2
1	10	3,8	100	38	2,91	0,89	0,80
2	20	4	400	80	3,36	0,64	0,41
3	40	4,2	1600	168	4,25	-0,05	0,00
4	60	4,8	3600	288	5,15	-0,35	0,12
5	90	5,5	8100	495	6,50	-1,00	0,99
6	110	6	12100	660	7,39	-1,39	1,94
7	130	7	16900	910	8,29	-1,29	1,66
8	140	8,1	19600	1134	8,74	-0,64	0,41
9	150	10	22500	1500	9,19	0,81	0,66
10	160	12	25600	1920	9,63	2,37	5,60
SUM	910	65,4	110500	7193	-	-	12,59

2. Составим и запишем систему уравнений для нахождения коэффициентов k и b

Вспомогательные величины	
Δ	276900
Δ1	12416
Δ2	681070

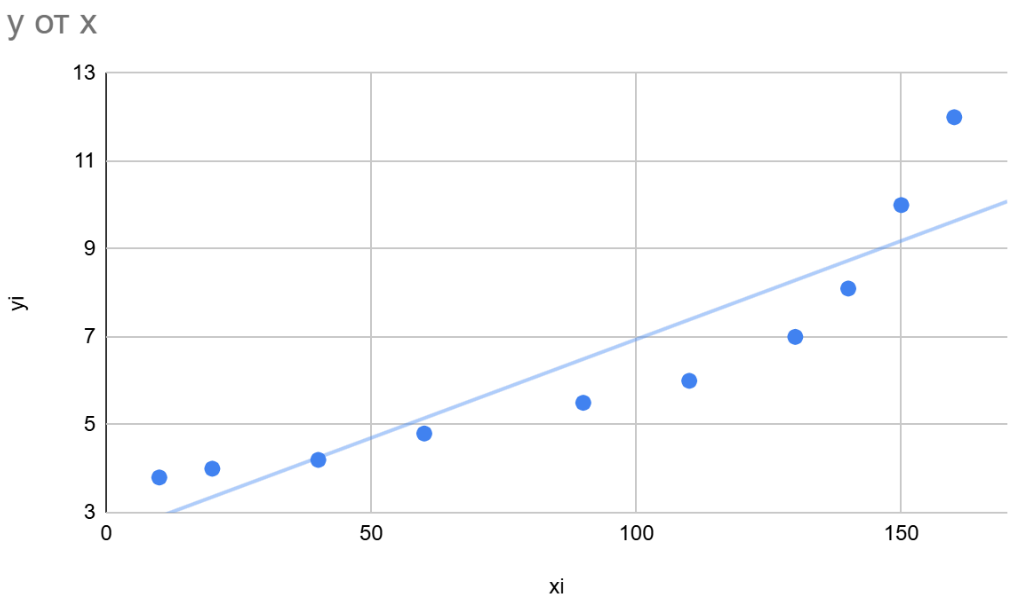
3. Неизвестные k и b найдем по формуле Крамера.

Вычисленные величины	
k	0,04
b	2,46

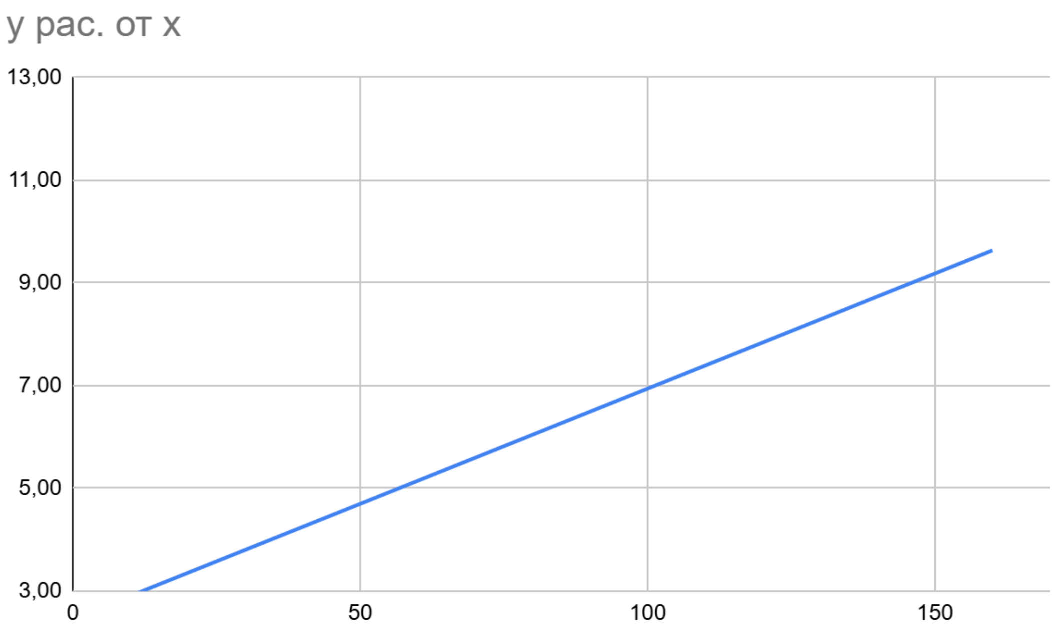
4. Составим и запишем уравнение y=kx+b

Вычисленный y при x = 170	
x	y
170	10,08

5. Построим график исходных данных.



6. Изобразим прямую регрессию на построенном графике. Построим график функции.



3.2 Задание 2

В таблице приведены данные о времени работы (t, у.е.) некоторого алгоритма в зависимости от количества его элементов (x).

x_i	9	12	14	16	18	20	21	23	24	25
t_i	152	280	380	500	630	780	860	1025	1130	1225

В предположении, что между x и y существует квадратич-

28

ная зависимость, определить параметры регрессии $t = a_2x^2 + a_1 + a_0$ методом наименьших квадратов. Спрогнозировать время работы алгоритма, состоящего из 30 элементов.

t = a2 * x^2 + a1 * x + a0

30 - ?

1. Введем исходные данные. Произведем необходимые вычисления.

№	xi	yi	x^2	x^3	x^4	x*y	x^2 * y	y рас
1	9	152	81	729	6561	1368	12312	154,16
2	12	280	144	1728	20736	3360	40320	277,67
3	14	380	196	2744	38416	5320	74480	379,67
4	16	500	256	4096	65536	8000	128000	497,42

5	18	630	324	5832	104976	11340	204120	630,91
6	20	780	400	8000	160000	15600	312000	780,14
7	21	860	441	9261	194481	18060	379260	860,65
8	23	1025	529	12167	279841	23575	542225	1 033,49
9	24	1130	576	13824	331776	27120	650880	1 125,81
10	25	1225	625	15625	390625	30625	765625	1 222,07
SUM	182	6962	3572	74006	1592948	144368	3109222	6 962,00

2. Составим и запишем систему уравнений для нахождения коэффициентов a2,a1,a0;

$$\begin{cases} \left(\sum_{i=1}^{n} x_i^4\right)\cdot a + \left(\sum_{i=1}^{n} x_i^3\right)\cdot b + \left(\sum_{i=1}^{n} x_i^2\right)c = \sum_{i=1}^{n} x_i^2 y_i \\ \left(\sum_{i=1}^{n} x_i^3\right)a + \left(\sum_{i=1}^{n} x_i^2\right)\cdot b + \left(\sum_{i=1}^{n} x_i\right)\cdot c = \sum_{i=1}^{n} x_i \cdot y_i \\ \left(\sum_{i=1}^{n} x_i^2\right)a + \left(\sum_{i=1}^{n} x_i\right)\cdot b + cn = \sum_{i=1}^{n} y_i \end{cases}$$

Вспомогательные величины	
Δ	13800648
Δ1	27152000
Δ2	-2059352
Δ3	-53203056

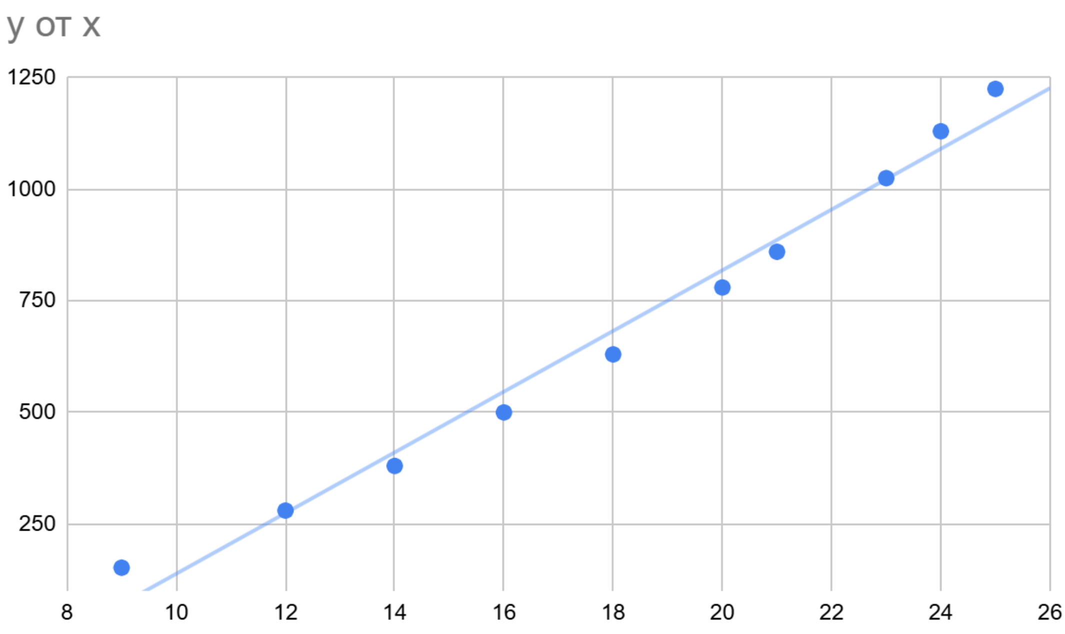
3. Найдем неизвестные коэффициенты a2,a1,a0

Вычисленные величины	
a2	1,97
a1	-0,15
a0	-3,86

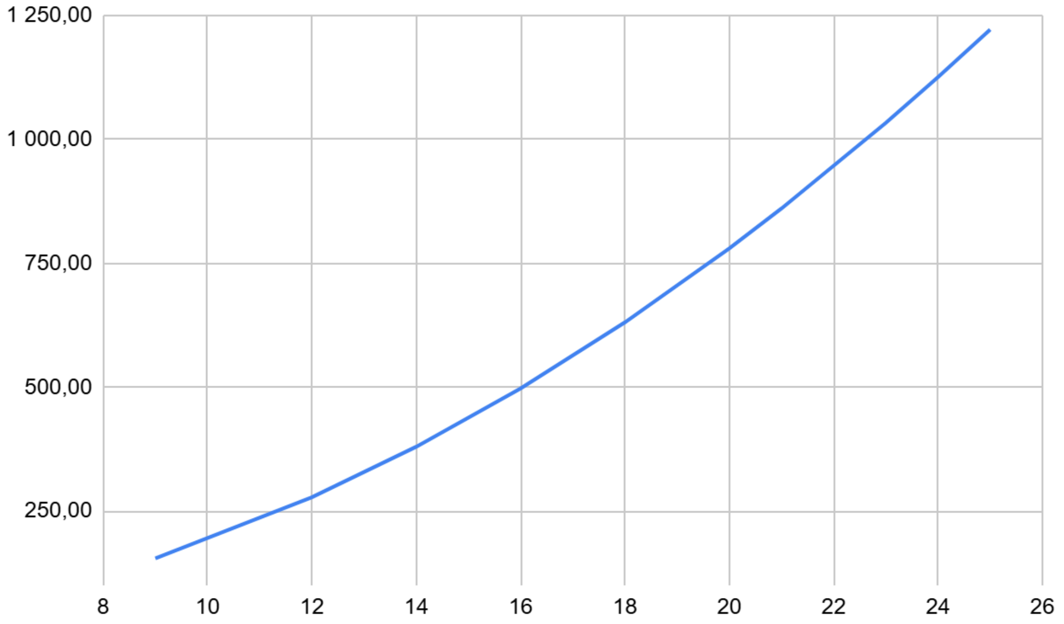
4. Составим и запишем уравнение. В рассматриваемом случае получаем уравнение:

Вычисленный y при x = 30	
x	y
30	1 762,37

5. Построим график исходных данных.По графику убедиться в возможной квадратичной зависимости между x и y.



6. Изобразим линию регрессию на построенном графике.



4. Вывод

В ходе работы были определены размах варьирования и составлен статистический ряд распределения частот. Также были построены полигон и гистограмма относительных частот, найдена эмпирическая функция распределения. Были вычислены выборочные значения математического ожидания, дисперсии и среднеквадратического отклонения. По виду полигона и гистограммы был выбран закон распределения. После этого были найдены точечные оценки параметров распределения и записана функция распределения и плотность вероятности. Наконец, была проведена проверка согласия эмпирической функции распределения с теоретической функцией при помощи критерия согласия Колмогорова.