ВВЕДЕНИЕ

Решение тех или иных вопросов, связанных с технологией добычи нефти и газа, базируется на промысловых материалах. В связи с этим, правильность принятых решений во многом зависит от способов обработки этих материалов и их анализа(планирование давлений гидроразрыва, дебита скважин, сроков службы глубинных насосов и т. д.)

Необходимо провести такое число наблюдений, опытов, чтобы установить, какой из законов распределения случайных величин действует в каждом конкретном случае. Если это удается, то можно предвидеть с некоторой вероятностью исход операции и составить конкретный план проведения мероприятия.

Однако установление функции распределения какого-либо параметра следует проводить достаточно осторожно. Иначе может быть так, что большое число проб, отобранных в скважине, в которой значения параметра оказались аномальными, значительно исказит действительное распределение параметра.

Таким образом, систематизация и обработка промыслового материала имеют большое значение.

Статистической совокупностью, состоящую из N единиц с конкретным значением количественного показателя может быть:

- дебит глубиннонасосной скважины;
- срок службы насосов;
- давление гидроразрыва;
- пластовое давление.

Например, при определении фактической подачи глубинных насосов диаметром 32 мм на месторождении получены следующие данные о дебите (в м3/сут): 2,3; 1,0; 0,4; 3,2; 2,5; 2,4; 1,3; 0,6; 0,5; 3,0; 1,6; 2,0; 1,4; 0,9; 2,2; 1,2; 0,7; 1,8; 0,6; 2,0; 2,0; 2,7; 2,0; 2,0; 2,5; 1,0; 1,8; 2,0; 1,2; 1,3.

I. ПЕРВИЧНАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ

Пусть X — некоторый производственный показатель (признак), а x_1, $x_2, \ldots, x_n$ — результаты независимых наблюдений над ними. Если количество наблюдений n невелико, наблюдения либо ранжируют, либо сводят в табл. 1, где каждому значению x_i ставят в соответствие частоту m_i появления этого значения в данной выборке.

Таблица 1

Варианты, x_i	x_1	x_2	...	x_n
частоты, m_i	m_1	m_2	...	m_n

Здесь $\sum_{i=1}^{k} m_i = n$, где n — объем выборки.

Замечание. Если количество наблюдений n достаточно большое ($n > 30$), то результаты наблюдений сводят в интервальный вариационный ряд.

АЛГОРИТМ ФОРМИРОВАНИЯ ВАРИАЦИОННОГО РЯДА

1. Вычисляют размах варьирования R признака X по формуле $\boxed{R = x_{\max} - x_{\min}}$
2. Полученный результат делится на k разных частей (выбираем по нижеприведенному правилу) и получаем число столбцов (интервалов) в таблице.

$$\boxed{1)\, 6 \le k \le 20\,, \qquad 2)\, k \approx \sqrt{n}\,, \qquad 3)\, k = 1 + 3{,}2\lg n\,.}$$

3. При небольшом объеме n выборки число интервалов принимают равным от 6 до 10. Длина h каждого частичного интервала определяется по формуле

$$h = \frac{R}{k}$$

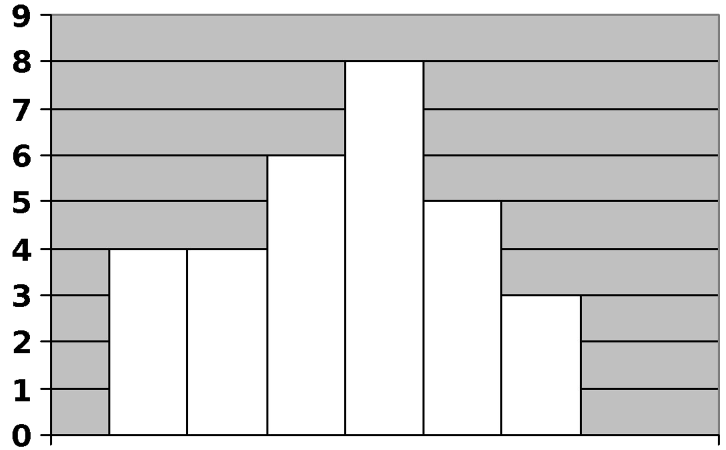
4. За начало x_0 первого интервала рекомендуется брать величину $x_0 = x_{\min} - 0{,}5h$. Конец x_k последнего интервала находят по формуле $x_k = x_{\max} + 0{,}5h$. Сформированный интервальный вариационный ряд записывают в виде табл. 2.

Таблица 2

Варианты-интервалы, $(x_i\,;\,x_{i+1})$	$(x_0\,;\,x_1)$	$(x_1\,;\,x_2)$	...	$(x_{k-1}\,;\,x_k)$
частоты, m_i	m_1	m_2	...	m_k

Интервальный вариационный ряд изображают геометрически в виде гистограммы частот m_i или гистограммы относительных частот $\frac{m_i}{n}$.

Гистограммой называется ступенчатая фигура, для построения которой по оси $0x$ откладывают отрезки, изображающие частичные интервалы $(x_i;x_{i+1})$ варьирования признака X, и на этих отрезках, как на основаниях, строят прямоугольники с высотами, равными частотам или частотам соответствующих интервалов.



ПЕРЕХОД ОТ ИНТЕРВАЛЬНОГО ВАРИАЦИОННОГО РЯДА К ДИСКРЕТНОМУ

В качестве вариантов x_i этого ряда берут середины интервалов $(x_i;\ x_{i+1})$. Дискретный вариационный ряд записывается в виде табл. 3 или табл. 4.

Таблица 3

Варианты, x_i	x_1	x_2	...	x_k
частоты, m_i	m_1	m_2	...	m_k

Здесь $\sum m_i = n$, где n — объем выборки.

Таблица 4

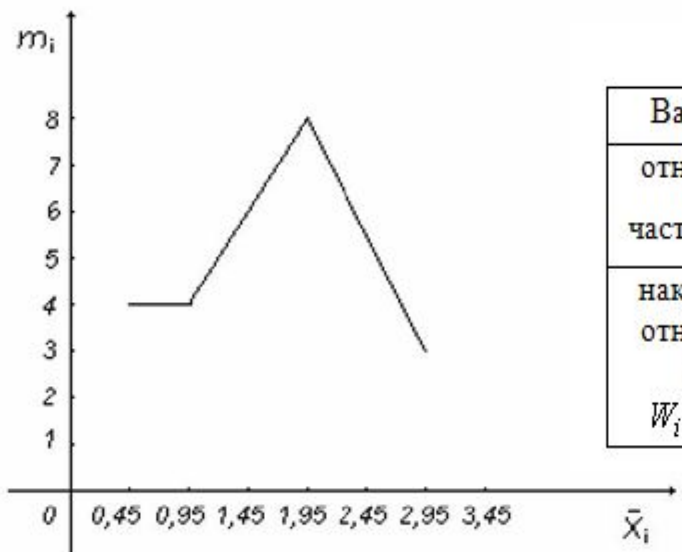
Варианты, x_i	x_1	x_2	...	x_k
относительные частоты, $\omega_i = \frac{m_i}{n}$	ω_1	ω_2	...	ω_k

Здесь $\sum_{i=1}^{k} \omega_i = 1$.

ИЗОБРАЖЕНИЕ

Полигон частот или относительных частот для дискретного вариационного ряда

В системе координат $0xy$ строят точки с координатами $(x_i\,;\,m_i)$ или $(x_i\,;\,\omega_i)$, где x_i — значение i-го варианта, а m_i (ω_i) — соответствующие частоты.



Полигон частот (эмпирическая кривая)

Кумулятивная кривая для вариационного ряда

При построении кумуляты дискретного вариационного ряда на оси абсцисс откладывают варианты x_i, а по оси ординат соответствующие им накопленные частоты W_i.
Соединяя точки с координатами $(x_i\,;\,W_i)$ отрезками прямых, получаем ломаную (кривую), которую называют кумулятой.

Таблица 5

Варианты, x_i	x_1	x_2	...	x_k
относительные частоты, $\omega_i = \frac{x_i}{n}$	$\omega_1 = \frac{x_1}{n}$	$\omega_2 = \frac{x_2}{n}$	...	$\omega_k = \frac{x_k}{n}$
накопительные относительные частоты, $W_i = W_{i-1} + \omega_i$	$W_1 = \omega_1$	$W_2 = W_1 + \omega_2$	...	$W_k = W_{k-1} + \omega_k$

ЭМПИРИЧЕСКАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ

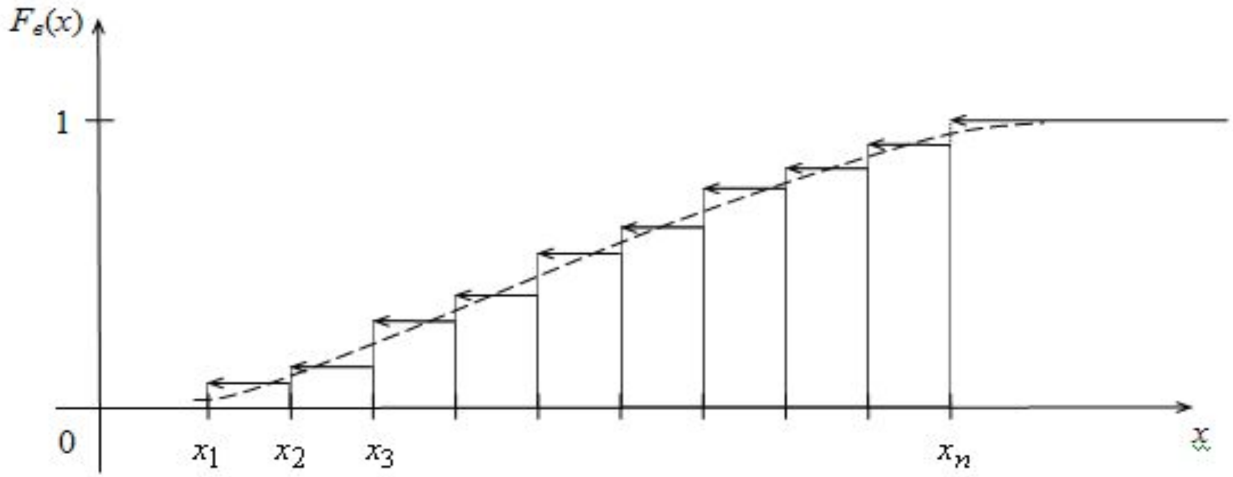
Эмпирической функцией распределения или функцией распределения называется функция $F_в(x)$, определяемая равенством

$$F_в(x) = \frac{m_x}{n}, \qquad (1)$$

где n — объем выборки,
m_x — число вариантов x_i, меньших x.

Эмпирическая функция $F_в(x)$ служит для оценки теоретической функции распределения генеральной совокупности.

Значения эмпирической функции $F_в(x)$ принадлежат промежутку $[0;1]$. Графиком функции $F_в(x)$ служит кусочно-постоянная кривая



II. РАСЧЕТ ВЫБОРОЧНЫХ ХАРАКТЕРИСТИК СТАТИСТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ

Для характеристики важнейших свойств статистического распределения используют средние показатели, называемые **выборочными числовыми характеристиками.**

Если значения x_i признака X не сгруппированы в вариационные ряды (табл. 2, 3, 4) и объем выборки n небольшой, то несмещенные оценки для неизвестных **математического ожидания a и дисперсии σ^2** находят по формулам:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (2)$$

для математического ожидания и

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} (x - \bar{x})^2 \qquad (3)$$

для дисперсии.

Если результаты наблюдений сгруппированы в дискретный вариационный ряд (табл. 3), то те же оценки находятся по формулам:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i m_i\,,\; n = \sum_{i=1}^{k} m_i\,, \qquad 4)$$

$$S^2 = \frac{1}{n}\sum_{i=1}^{k} (x_i - \bar{x})^2 m_i\,, \qquad (5)$$

По формуле (5) вычисляют S^2 в случае, если объем выборки $n \geq 50$.

Если же $n < 50$, то вычисляют исправленную дисперсию $\tilde{S}^2$ по формуле:

$$\tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 \tag{6}$$

для простой выборки или

$$\tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 m_i \tag{7}$$

для взвешенной выборки.

Выборочное среднее квадратичное отклонение находят по формулам

$$S = \sqrt{S^2} \tag{8}$$

при различных объемах выборки.

ХАРАКТЕРИСТИКИ

Модой M_0X называют варианту, которая имеет наибольшую частоту. Например, для вариационного ряда

x_i	4	9	14	19
m_i	3	7	2	5

$M_0X = 9$.

Медианой M_lX называют варианту, которая делит вариационный ряд на равные части.

При нечетном объеме выборки $n = 2k + 1$ медиана равна серединному члену вариационного ряда.

Например, для вариационного ряда

x_i	3	5	**8**	12	15
m_i	6	2	4	5	8

$M_lX = 8$.

При четном объеме выборки $n = 2k$ медиана находится по формуле

$$M_lX = \frac{x_k + x_{k+1}}{2} \qquad (9)$$

x_i	2	5	**7**	**10**	12	14
m_i	3	4	8	2	3	6

$M_lX = \frac{7+10}{2} = 8{,}5$.

Пользуясь расчетной табл. 6. можно вычислить

начальные моменты (при достаточно большом объеме выборки $(n > 30)$)

Таблица 6

x_i	m_i	u_i	$m_i u_i$	$m_i u_i^2$	$m_i u_i^3$	$m_i u_i^4$	контрольный столбец $m_i(u_i+1)^2$
строка сумм:	$\sum =$		$\sum =$	$\sum =$	$\sum =$	$\sum =$	$\sum =$

$u_i = \frac{x_i - C}{h}$ условные варианты

$C = M_0 X$, h — шаг (длина интервала).

Контроль вычислений ведут по формуле

$$\sum m_i + 2\sum m_i u_i + \sum m_i u_i^2 = \sum m_i (u_i + 1)^2 . \qquad (10)$$

Начальные моменты

$$M_1^* = \frac{1}{n}\sum m_i u_i \qquad (11)$$

$$M_2^* = \frac{1}{n}\sum m_i u_i^2 \qquad (12)$$

$$M_3^* = \frac{1}{n}\sum m_i u_i^3 \qquad (13)$$

$$M_4^* = \frac{1}{n}\sum m_i u_i^4 \qquad (14)$$

Выборочная средняя

$$\bar{x} = M_1^* h + C \qquad (15)$$

Выборочная дисперсия

$$S^2 = \left(M_2^* - \left(M_1^*\right)^2\right)h^2 \qquad (16)$$

Выборочное среднее квадратическое отклонение

$$S = \sqrt{S^2} \qquad (17)$$

Асимметрия и эксцесс

$$A_s = \frac{m_3}{S^3} \qquad (18)$$

$$E_x = \frac{m_4}{S^4} - 3, \text{ где}$$

условный центральный момент третьего порядка

(19)

$$m_3 = \left(M_3^* - 3M_2^* M_1^* + 2\left(M_1^*\right)^3\right)h^3$$

(20)

условный центральный момент четвертого порядка.

$$m_4 = \left(M_4^* - 4M_3^* M_1^* + 6M_2^*\left(M_1^*\right)^2 - 3\left(M_1^*\right)^4\right)h^4 \qquad (21)$$

Для характеристики колеблемости признака X используют относительный показатель — **коэффициент вариации V**, который вычисляют по формуле

$$V = \frac{S}{\bar{x}} \cdot 100\% \qquad (22)$$

III. ИНТЕРВАЛЬНЫЕ (ДОВЕРИТЕЛЬНЫЕ) ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

При ограниченных объемах выборки возникает необходимость указать степень точности и надежности оценок генеральных характеристик.

При решении практических задач, как правило, значения генеральной дисперсии и математического ожидания неизвестны.

Для оценки генеральной средней, то есть $M(X)=a$ и генерального среднеквадратического отклонения σ по выборочной средней $\bar{x}$ и выборочному среднеквадратическому отклонению S находят **доверительные интервалы** по формулам

$$\bar{x}-\frac{S}{\sqrt{n}}\cdot t_{\gamma}<a<\bar{x}+\frac{S}{\sqrt{n}}\cdot t_{\gamma}, \qquad (23)$$

где t_{γ} находят **по таблице распределения Стьюдента** по заданным n и γ (γ – уровень доверия или надежность, которая задается заранее).

Для генерального среднеквадратического отклонения доверительные интервалы находят по формулам

$$S(1-q) < \sigma < S(1+q),\ \text{при}\ (q < 1) \qquad (24)$$

или

$$0 < \sigma < S(1+q),\ \text{при}\ (q > 1) \qquad (25)$$

величину q находят по таблице значений $q = (\gamma, n)$ (приложение учебника по теории вероятностей и математической статистике) по заданным n и γ.

ЛАБОРАТОРНАЯ РАБОТА № 1

Построение вариационных рядов. Расчет числовых характеристик.

Цель работы: овладение студентом способами построения рядов распределения и методами расчета числовых характеристик.

Содержание работы: на основе совокупности данных опыта выполнить следующее:

1. Построить ряды распределения (интервальный и дискретный вариационные ряды). Изобразить их графики.
2. Построить график накопительных частот — кумуляту.
3. Составить эмпирическую функцию распределения и изобразить ее графически.
4. Вычислить моду, медиану, выборочную среднюю, выборочную дисперсию, выборочное среднее квадратическое отклонение, коэффициент вариации, асимметрию, эксцесс.
5. Построить доверительные интервалы для истинного значения измеряемой величины и среднего квадратического отклонения генеральной совокупности.
6. Раскрыть смысловую сторону каждой характеристики.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ РАБОТЫ

1. Построить интервальный вариационный ряд. Для этого найти:

а) Размах варьирования признака по формуле $R = x_{\max} - x_{\min}$, где $x_{\min}$ — наименьшая варианта, $x_{\max}$ — наибольшая варианта в данной выборочной совокупности;

б) число интервалов вариационного ряда, пользуясь одним из приведенных ниже соотношений:

$S \approx \sqrt{n}$, $6 < S < 12$, $S = 1 + 3{,}2\lg n$, где n — объем выборки;

в) длину частичных интервалов по формуле $h = \frac{R}{S}$ и по необходимости округлить это значение до некоторого числа.

Записать полученный вариационный ряд, заполнив табл. 1:

Варианты-интервалы, $(x_{i-1};\ x_i)$	$(x_0;\ x_1)$	$(x_1;\ x_2)$	...	$(x_{k-1};\ x_k)$
частоты, m_i	m_1	m_2	...	m_k

2. Построить дискретный вариационный ряд, взяв в качестве вариантов середины вариантов-интервалов непрерывного вариационного ряда, а в качестве частот частоты непрерывного вариационного ряда.

3. Изобразить графически интервальный и дискретный вариационные ряды (построить гистограмму и полигон частот).

4. Построить график накопленных частот — кумуляту. Кумулята — это ломаная линия, проходящая через точки с координатами x_i и соответствующими накопленными частотами. Предварительно составить табл. 2:

Таблица 2

Варианты, x_i	x_1	x_2	...	x_k
относительные частоты, $\omega_i = \frac{x_i}{n}$	$\omega_1 = \frac{x_1}{n}$	$\omega_2 = \frac{x_2}{n}$	...	$\omega_k = \frac{x_k}{n}$
накопительные относительные частоты, $W_i = W_{i-1} + \omega_i$	$W_1 = \omega_1$	$W_2 = \omega_1 + \frac{x_2}{n}$	...	

5. Найти эмпирическую функцию распределения и изобразить ее графически.

6. Найти моду M_0X и медиану M_lX.

7. Для вычисления остальных статистик воспользоваться методом произведений. Ввести условные варианты $u_i = \frac{x_i - C}{h}$, где $C = M_0X$, h — шаг (длина интервала). Составить расчетную табл. 3:

Т а б л и ц а 3

x_i	m_i	u_i	m_iu_i	$m_iu_i^2$	$m_iu_i^3$	$m_iu_i^4$	контрольный столбец $m_i(u_i+1)^2$
строка сумм:	$\sum=$		$\sum=$	$\sum=$	$\sum=$	$\sum=$	$\sum=$

Контроль вычислений произвести по формуле

$$\sum m_i + 2\sum m_iu_i + \sum m_iu_i^2 = \sum m_i(u_i+1)^2 .$$

8. Пользуясь табл. 3, вычислить начальные моменты по формулам:

$$M_1^* = \frac{1}{n}\sum m_iu_i ,\ M_2^* = \frac{1}{n}\sum m_iu_i^2 ,\ M_3^* = \frac{1}{n}\sum m_iu_i^3 ,\ M_4^* = \frac{1}{n}\sum m_iu_i^4 .$$

9. Найти выборочную среднюю по формуле $\bar{x} = M_1^*h + C$.

10. Найти выборочную дисперсию по формуле $D(X) = (M_2^* - (M_1^*)^2)h^2$.

11. Найти выборочное среднее квадратическое отклонение по формуле $\sigma_в = S_X = \sqrt{D(X)}$.

12. Найти коэффициент вариации по формуле $V = \frac{S}{\bar{x}} \cdot 100\%$.

13. Вычислить асимметрию и эксцесс

14. Доверительные интервалы для «a» и «σ» найти по формулам

$$\bar{x} - t_\gamma \cdot \frac{S_x}{\sqrt{n}} < a < \bar{x} + t_\gamma \cdot \frac{S_x}{\sqrt{n}}, \text{ где } \gamma = 0{,}95 .$$

(t_γ найти по приложению учебника)

$$S_X(1-q) < \sigma < S_X(1+q), \text{ где } S_X = \sqrt{S_X^2} \text{ и } q < 1$$

$$0 < \sigma < S_X(1+q), \text{ при } q > 1.$$

(q найти по приложению учебника)

15. Раскрыть смысловую сторону каждой характеристики.

ДАННЫЕ ОБ ОБВОДНЕННОСТИ НЕФТИ ИЗ НАСОСНЫХ СКВАЖИН (в %)

61,2	61,4	60,2	61,2	61,3	60,4	61,4	60,8	61,2	60,6
61,6	60,2	61,3	60,3	60,7	60,9	61,2	60,5	61,0	61,4
61,1	60,9	61,5	61,4	60,6	61,2	60,1	61,3	61,1	61,3
60,3	61,3	60,6	61,7	60,6	61,2	60,8	61,3	61,0	61,2
60,5	61,4	60,7	61,3	60,9	61,2	61,1	61,3	60,9	61,4
60,7	61,2	60,3	61,1	61,0	61,5	61,3	61,9	61,4	61,3
61,6	61,0	61,7	61,1	60,9	61,5	61,6	61,4	61,5	61,2
61,6	61,3	61,8	61,1	61,7	60,9	62,2	61,1	62,1	61,0
61,5	61,7	62,3	62,3	61,7	62,9	62,5	62,8	62,6	61,5
62,1	62,6	61,6	62,5	62,4	62,3	62,1	62,3	62,2	62,1

Пусть ***X*** – обводненность нефти из рассматриваемых насосных скважин

1. СТРОИМ ИНТЕРВАЛЬНЫЙ ВАРИАЦИОННЫЙ РЯД

$x_{max} = 62,8$, $x_{min} = 60,1$. Тогда размах варьирования признака $R = 62,8 - 60,1 = 2,7$. Определяем число интервалов (число столбцов в таблице) вариационного ряда. Пусть $S = 10$. Длину каждого частичного интервала определяем по формуле $h = \frac{R}{S} = \frac{2,7}{10} = 0,27$. Так как исходные данные мало отличаются друг от друга и содержат один десятичный знак, то величину h округляем до одного десятичного знака, то есть берем $h = 0,3$.

Подсчитываем число вариантов, попадающих в каждый интервал, по данным выборки. Значение x_i, попадающее на границу интервала, относим к левому концу. За начало x_0 первого интервала берем величину $\boxed{x_0 = x_{\min} - 0{,}5h = 60{,}1 - 0{,}5 \cdot 0{,}3 = 59{,}95 \approx 60}$. Конец x_k последнего интервала находим по формуле $\boxed{x_k = x_{\max} + 0{,}5h = 62{,}8 + 0{,}15 = 62{,}95 \approx 63{,}0}$. Сформированный интервальный вариационный ряд записываем в виде табл. 1.

Таблица 7

Варианты-интервалы	60- 60,3	60,3- 60,6	60,6- 60,9	60,9- 61,2	61,2- 61,5	61,5- 61,8	61,8- 62,1	62,1- 62,4	62,4- 62,7	62,7- 63,0
Частоты, m_i	3	6	9	18	29	16	2	11	5	1

Контроль: $\sum m_i = 100$ и объем выборки $n = 100$.

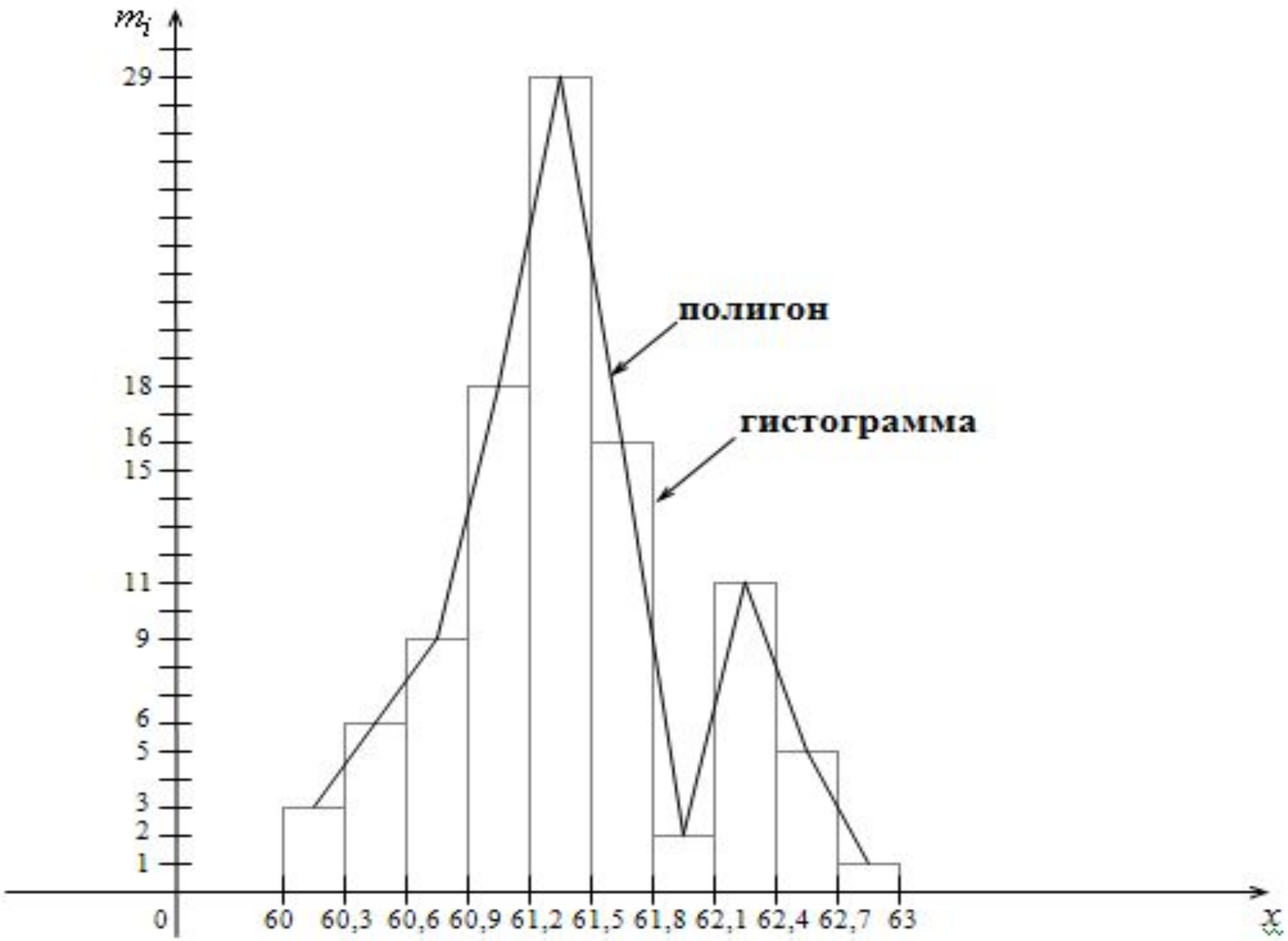
2. СТРОИМ ДИСКРЕТНЫЙ ВАРИАЦИОННЫЙ РЯД

В качестве вариантов x_i берем середины интервалов интервального вариационного ряда.

Таблица 8

варианты, x_i	60,15	60,45	60,75	61,05	61,35	61,65	61,95	62,25	62,55	62,85
частоты, m_i	3	6	9	18	29	16	2	11	5	1

3. ГРАФИЧЕСКОЕ ИЗОБРАЖЕНИЕ ИНТЕРВАЛЬНОГО И ДИСКРЕТНОГО ВАРИАЦИОННОГО РЯДА

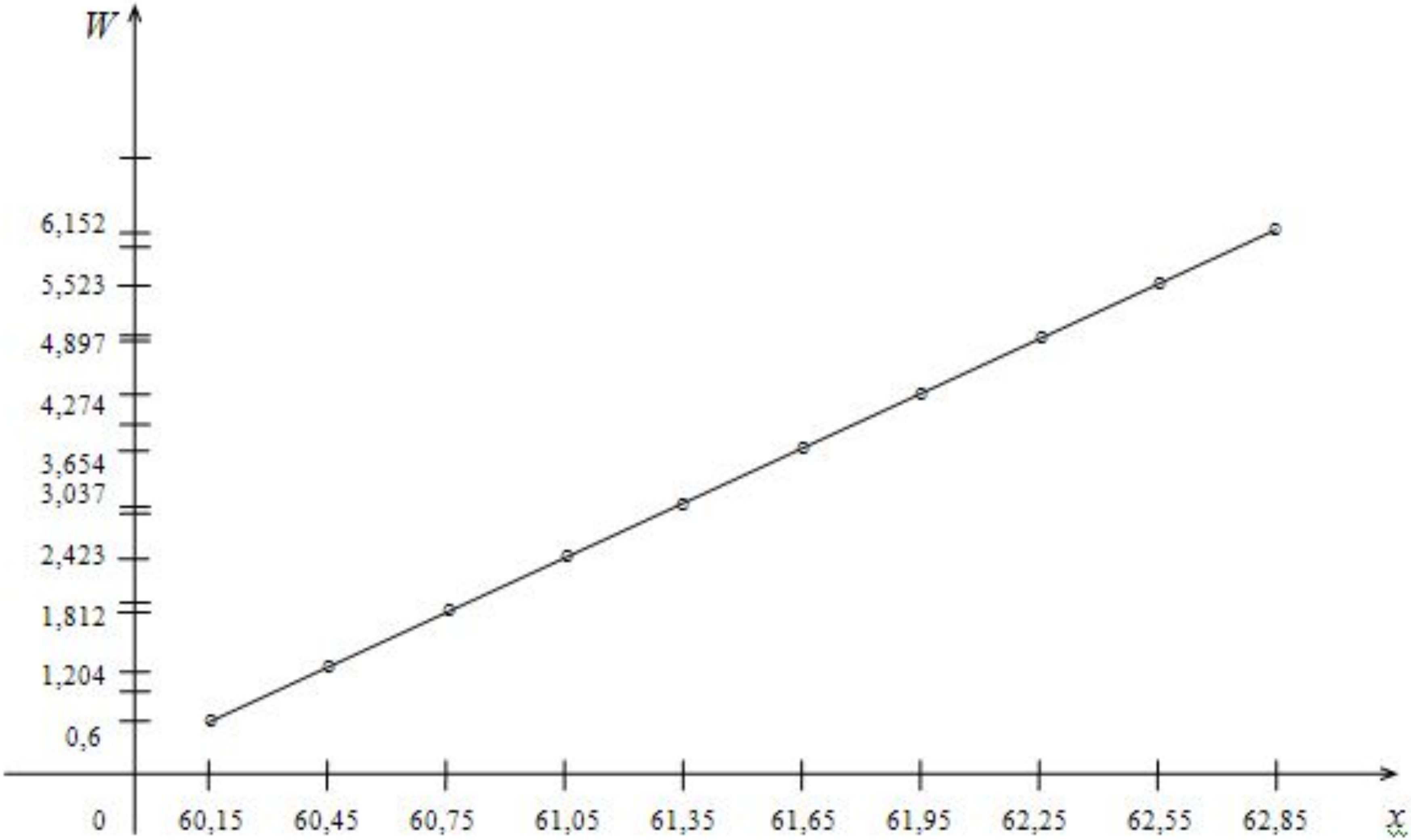


<u>4. СТРОИМ КУМУЛЯТУ</u>

Предварительно составляем расчетную таблицу.

Таблица 9

Варианты, x_i	60,15	60,45	60,75	61,05	61,35	61,65	61,95	62,25	62,55	62,85
относительные частоты, $\omega_i = \frac{x_i}{n}$	0,602	0,604	0,608	0,611	0,614	0,617	0,620	0,623	0,626	0,629
накопительные относительные частоты, $W_i = W_{i-1} + \omega_i$	0,6	1,204	1,812	2,423	3,037	3,654	4,274	4,897	5,523	6,152



W
6,152
5,523
4,897
4,274
3,654
3,037
2,423
1,812
1,204
0,6
0
60,15
60,45
60,75
61,05
61,35
61,65
61,95
62,25
62,55
62,85
x

5. НАХОДИМ ЭМПИРИЧЕСКУЮ ФУНКЦИЮ РАСПРЕДЕЛЕНИЯ

$F_в(x) = \frac{m_x}{n}$, где n — объем выборки, m_x — накопленные частоты.

При $x \le 60{,}15$ $F_B(x) = 0$

При $60{,}5 < x \le 60{,}45$ $F_в(x) = \frac{m_x}{n} = \frac{3}{100} = 0{,}03$.

При $60{,}45 < x \le 60{,}75$ $F_в(x) = \frac{m_x}{n} = \frac{3+6}{100} = 0{,}09$.

При $60{,}75 < x \le 61{,}05$ $F_в(x) = \frac{m_x}{n} = \frac{3+6+9}{100} = 0{,}18$.

При $61{,}05 < x \le 61{,}35$ $F_в(x) = \frac{m_x}{n} = \frac{18+18}{100} = 0{,}36$.

При $61{,}35 < x \le 61{,}65$ $F_в(x) = \frac{m_x}{n} = \frac{36+29}{100} = 0{,}65$.

При $61{,}65 < x \le 61{,}95$ $F_в(x) = \frac{m_x}{n} = \frac{65+16}{100} = 0{,}81$.

При $61{,}95 < x \le 62{,}25$ $F_в(x) = \frac{m_x}{n} = \frac{81+2}{100} = 0{,}83$.

При $62{,}25 < x \le 62{,}55$ $F_в(x) = \frac{m_x}{n} = \frac{83+11}{100} = 0{,}94$.

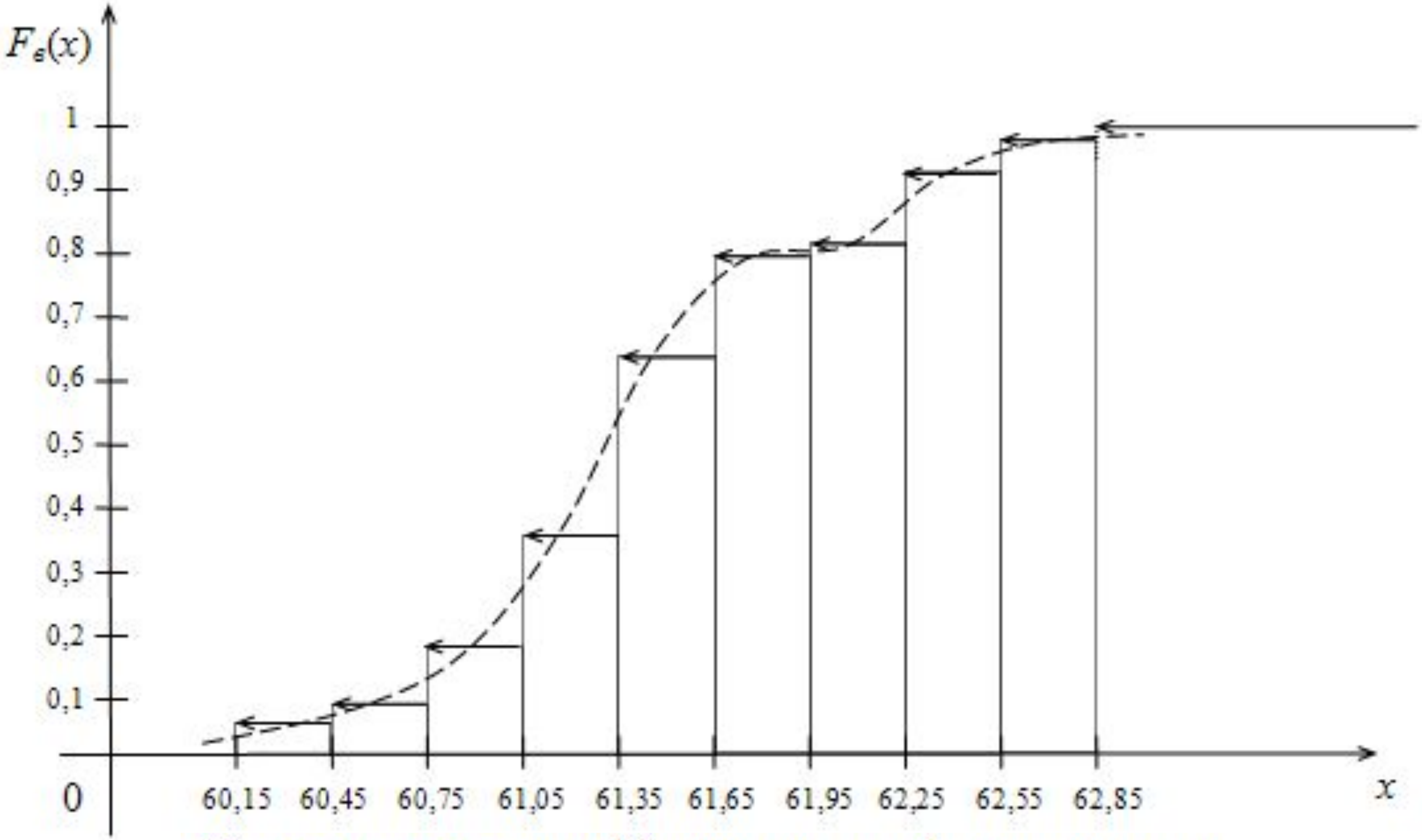
При $62{,}55 < x \le 62{,}85$ $F_в(x) = \frac{m_x}{n} = \frac{94+5}{100} = 0{,}99$.

При $x > 62{,}85$ $F_в(x) = 1$

Записываем полученную эмпирическую функцию в виде:

$$F_в(x) = \begin{cases} 0 \quad \text{при}\, x \le 60{,}15, \\ 0{,}03 \quad \text{при} \;\; 60{,}15 < x \le 60{,}45, \\ 0{,}09 \quad \text{при} \;\; 60{,}45 < x \le 60{,}75, \\ 0{,}18 \quad \text{при} \;\; 60{,}75 < x \le 61{,}05, \\ 0{,}36 \quad \text{при} \;\; 61{,}05 < x \le 61{,}35, \\ 0{,}65 \quad \text{при} \;\; 61{,}35 < x \le 61{,}65, \\ 0{,}81 \quad \text{при} \;\; 61{,}65 < x \le 61{,}95, \\ 0{,}83 \quad \text{при} \;\; 61{,}95 < x \le 62{,}25, \\ 0{,}94 \quad \text{при} \;\; 62{,}25 < x \le 62{,}55, \\ 0{,}99 \quad \text{при} \; 62{,}55 < x \le 62{,}85, \\ 1 \quad \text{при} \;\; x > 62{,}85. \end{cases}$$

График функции $F_в(x)$ имеет вид:



Абсциссами точек кривой (штриховая линия) служат значения обводненности нефти, добываемой насосным способом из скважин.

Ординатами — значения эмпирической функции распределения, т.е. вероятности попадания возможных значений обводненности нефти на промежуток $(-\infty; x_i]$.

6. МОДА И МЕДИАНА (см. табл. № 8)

$M_0 X = 61{,}35$, то есть это значение обводненности нефти, встречающееся в данной выборке с наибольшей частотой ($m = 29$).

Так как табл. 8 содержит четное число столбцов, то $M_l X = \frac{61{,}35 + 61{,}65}{2} = 61{,}5$. Это значение обводненности нефти, которое делит данные выборки признака X на равные части.

7. МЕТОД ПРОИЗВЕДЕНИЙ ДЛЯ НАХОЖДЕНИЯ ОСТАЛЬНЫХ СТАТИСТИК

Введем условные варианты $u_i = \frac{x_i - C}{h}$.

У нас $C = M_0X = 61{,}35$, $h = 0{,}3$.

Составим расчетную табл. 10.

Таблица 10

x_i	m_i	u_i	m_iu_i	$m_iu_i^2$	$m_iu_i^3$	$m_iu_i^4$	контрольный столбец $m_i(u_i+1)^2$
60,15	3	-4	-12	48	-192	768	27
60,45	6	-3	-18	54	-162	486	24
60,75	9	-2	-18	36	-72	144	9
61,05	18	-1	-18	18	-18	18	0
61,35	29	0	0	0	0	0	29
61,65	16	1	16	16	16	16	64
61,95	2	2	4	8	16	32	18
62,25	11	3	33	99	297	891	176
62,55	5	4	20	80	320	1280	125
62,85	1	5	5	25	125	625	36
	100		12	384	330	4260	508

Контроль вычислений проводим по формуле

$$\sum m_i + 2\sum m_iu_i + \sum m_iu_i^2 = \sum m_i(u_i+1)^2 .$$

У нас $100 + 2 \cdot 12 + 384 = 508$, $508 = 508$.

Следовательно, вычисления проведены верно.

8. ВЫЧИСЛИМ НАЧАЛЬНЫЕ МОМЕНТЫ (см. табл. № 10)

$$M_1^* = \frac{1}{n}\sum m_i u_i = \frac{1}{100}\cdot 12 = 0{,}12\,.$$

$$M_2^* = \frac{1}{n}\sum m_i u_i^2 = \frac{1}{100}\cdot 384 = 3{,}84\,.$$

$$M_3^* = \frac{1}{n}\sum m_i u_i^3 = \frac{1}{100}\cdot 330 = 3{,}3\,.$$

$$M_4^* = \frac{1}{n}\sum m_i u_i^4 = \frac{1}{100}\cdot 4260 = 42{,}6\,.$$

9. ВЫБОРОЧНАЯ СРЕДНЯЯ

$$\bar{x} = M_1^* h + C = 0{,}12 \cdot 0{,}3 + 61{,}35 = 61{,}385 \approx 61{,}38 .$$

$\bar{x} = 61{,}38$ характеризует среднюю обводненность нефти из насосных скважин в данной выборке, составляющую 61,38%.

10. ВЫБОРОЧНАЯ ДИСПЕРСИЯ

$$S^2 = (M_2^* - (M_1^*)^2) h^2 = (3{,}84 - 0{,}12^2) \cdot 0{,}09 = 0{,}3443 .$$

11. ВЫБОРОЧНОЕ СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ

$$S = \sqrt{S^2} = \sqrt{0{,}3443} = 0{,}59 .$$

Величина $S = 0{,}59$ характеризует степень рассеяния значений обводненности нефти относительно средней обводненности.

12. КОЭФФИЦИЕНТ ВАРИАЦИИ

$$V = \frac{S}{\bar{x}} \cdot 100\% = \frac{0,59}{61,38} \cdot 100\% = 0,96\% .$$

Показывает тесную сгруппированность значений обводненности нефти около центра рассеяния, т.е. около средней обводненности нефти.

13. АСИММЕТРИЯ И ЭКСЦЕСС

Сначала находим центральные моменты третьего и четвертого порядков по формулам:

$$m_3 = (M_3^* - 3M_2^* M_1^* + 2(M_1^*)^3)h^3 = (3{,}3 - 3 \cdot 3{,}84 \cdot 0{,}12 + 2 \cdot 0{,}12^3) \cdot 0{,}3^3 = 0{,}05.$$

$$m_4 = (M_4^* - 4M_3^* M_1^* + 6M_2^*(M_1^*)^2 - 3(M_1^*)^4)h^4 = (42{,}6 - 4 \cdot 3{,}3 \cdot 0{,}12 + 6 \cdot 3{,}84 \cdot 0{,}12^2 - 3 \cdot 0{,}12^4) \cdot 0{,}3^4 = 0{,}33.$$

Тогда

$$A_s = \frac{m_3}{S^3} = \frac{0{,}05}{0{,}59^3} = 0{,}24.$$

$$E_x = \frac{m_4}{S^4} - 3 = \frac{0{,}33}{0{,}59^4} - 3 = -0{,}28.$$

ВЫВОД: значения A_s и E_x мало отличаются от нуля. Поэтому можно предположить близость данной выборки, характеризующей обводненность нефти, к нормальному распределению.

14. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ ДЛЯ «*a*» и «σ»

Произведем оценку **генеральной средней** $M(X) = a$ и **генерального среднеквадратического отклонения** $\sigma = S$ по выборочным статистикам $\bar{x}$ и S, используя теорию доверительных интервалов для нормального распределения.

Доверительный интервал для истинного значения обводненности нефти находим с надежностью $\gamma = 0{,}95$ по формуле

$$\bar{x} - \frac{S}{\sqrt{n}} \cdot t_\gamma < a < \bar{x} + \frac{S}{\sqrt{n}} \cdot t_\gamma .$$

По таблице приложения № 3 при $n = 100$ и $\gamma = 0{,}95$ находим $t_\gamma = 1{,}984$.

Записываем доверительный интервал

$$61{,}38 - \frac{0{,}59}{10} \cdot 1{,}984 < a < 61{,}38 + \frac{0{,}59}{10} \cdot 1{,}984$$

или $61{,}26 < a < 61{,}50$.

Запишем **доверительный интервал**
для генерального среднеквадратического отклонения
$\sigma = S$.

При заданных $\gamma = 0{,}95$ и $n = 100$ по таблице приложения № 4 находим $q = 0{,}143$.

Так как $q < 1$, то доверительный интервал записываем в виде

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

или $0{,}59(1 - 0{,}143) < \sigma < 0{,}59(1 + 0{,}143)$

или $0{,}50 < \sigma < 0{,}67$

ВЫВОД:

1. средняя обводненность нефти из насосных станций (в %) по данным выборки должна находиться в промежутке $(61{,}26; 61{,}50)$.

2. отклонения истинных значений обводненности нефти из насосных станций не должны выходить за пределы промежутка $(0{,}50; 0{,}67)$.